

Büchi automata recognizing sets of reals definable in first-order logic with addition and order

Arthur Milchior

November 14, 2016

Abstract

This work considers weak deterministic Büchi automata reading encodings of non-negative reals in a fixed base. A Real Number Automaton is an automaton which recognizes all encoding of elements of a set of reals. It is explained how to decide in linear time whether a set of reals recognized by a given minimal weak deterministic RNA is $\text{FO}[\mathbb{R}; +, <, 1]$ -definable. Furthermore, it is explained how to compute in quasi-quadratic (respectively, quasi-linear) time an existential (respectively, existential-universal) $\text{FO}[\mathbb{R}; +, <, 1]$ -formula which defines the set of reals recognized by the automaton. It is also shown that techniques given by Muchnik and by Honkala for automata over vector of natural numbers also works on vector of real numbers. It implies that some problems such as deciding whether a set of tuples of reals $R \subseteq \mathbb{R}^d$ is a subsemigroup of $(\mathbb{R}^d, +)$ or is $\text{FO}[\mathbb{R}; +, <, 1]$ -definable is decidable.

Introduction

This paper deals with logically defined sets of numbers encoded by weak deterministic Büchi automata. The sets of tuples of integers whose encodings in base b are recognized by a finite automaton are called the b -recognizable sets. By [5], the b -recognizable sets of vectors of integers are exactly the sets which are $\text{FO}[\mathbb{Z}; +, <, V_b]$ -definable, where $V_b(n)$ is the greatest power of b dividing n . It was proven in [15, 6] that the $\text{FO}[\mathbb{N}; +]$ -definable sets are exactly the sets which are b - and b' -recognizable for every $b \geq 2$.

The preceding results naturally led to the following problem: deciding whether a finite automaton recognizes a $\text{FO}[\mathbb{N}; +]$ -definable set of d -tuples of integers for some dimension $d \in \mathbb{N}^{>0}$. In the case of dimension $d = 1$, the decidability was proven in [9]. For $d > 1$, the decidability was proven in [14]. Another algorithm was given in [11], which solves this problem in polynomial time. For $d = 1$, a quasi linear time algorithm was given in [13].

The above-mentioned results about sets of tuples of natural numbers and finite automata have then been extended to results about set of tuples of reals recognized by a Büchi automata. The notion of Büchi automata is a formalism which describes languages of infinite words, also called ω -words. The Büchi automata are similar to the finite automata. The main difference between the two kinds of automata is that finite automata accept finite words which admits runs ending on accepting state, while Büchi automata accepts infinite words which admit runs in which an accepting state appears infinitely often.

One of the main differences between finite automata and Büchi automata is that finite automata can be determinized while deterministic Büchi automata are less expressive than Büchi automata. For example, the language $L_{\text{fin } a}$ of words containing a finite number of times the letter a is recognized by a Büchi automaton, but is not recognized by any deterministic Büchi automaton. This statement implies, for example, that no deterministic Büchi automaton recognizes the set of reals of the form nb^p with $n \in \mathbb{N}$ and $p \in \mathbb{Z}$, that is, the reals which admits no encoding in base b with a finite number of non-0 digits.

Another main difference between the two classes of automata is that the class of languages recognized by finite automata is closed under complement while the class of languages recognized by deterministic Büchi automata is not closed under complement. For example, $L_{\text{inf } a}$, the complement of $L_{\text{fin } a}$, is recognized by a deterministic Büchi automaton.

A Real Vector Automaton (RVA, See e.g. [3]) of dimension d is a Büchi automaton \mathcal{A} of alphabet $\{0, \dots, b-1\}^d \cup \{\star\}$, which recognizes the set of encoding in base b of the elements of a set of vectors of reals. Equivalently, for w an infinite word encoding a vector of dimension d of real (r_0, \dots, r_{d-1}) , if w is recognized by \mathcal{A} , then all encodings w' of (r_0, \dots, r_{d-1}) are recognized by \mathcal{A} . In the case where the dimension d is 1, those automata are called Real Number Automata (RNA, See e.g. [2]).

The sets of tuples of reals whose encoding in base b is recognized by a RVA are called the b -recognizable sets. By [18], they are exactly the $\text{FO}[\mathbb{R}, \mathbb{Z}; +, <, X_b, 1]$ -definable sets. The logic $\text{FO}[\mathbb{R}, \mathbb{Z}; +, <, X_b, 1]$ is the first-order logic over reals with a unary predicate which holds over integers, addition, order, the constant one, and the function $X_b(x, u, k)$. The function $X_b(x, u, k)$ holds if and only if u is equal to some b^n with $n \in \mathbb{Z}$ and there exists an encoding in base b of x whose digit in position n is k . That is, u and x are of the form:

$$\begin{array}{cccccccccccc} u = & 0 & \dots & 0 & \star & 0 & \dots & 0 & 1 & 0 & \dots \\ x = & & & & \star & & & & k & & \dots \end{array}$$

or of the form:

$$\begin{array}{cccccccccccc} u = & 0 & \dots & 0 & 1 & 0 & \dots & 0 & \star & 0 & \dots \\ x = & & & & k & & & & \star & & \dots \end{array}$$

A weak deterministic Büchi automaton is a deterministic Büchi automaton whose set of accepting states is a union of strongly connected components. A set is said to be weakly b -recognizable if it is recognized by a weak automaton in

base b . By [3], a set is $\text{FO}[\mathbb{R}, \mathbb{Z}; +, <]$ -definable if and only if its set of encodings is weakly b -recognizable for all $b \geq 2$. The class of weak deterministic Büchi automata is less expressive than the class of deterministic Büchi automata. For example, the language $L_{\inf a}$ of words containing an infinite number of a is recognized by a deterministic Büchi automaton but is not recognized by any weak deterministic Büchi automaton. This implies that, for example, no weak deterministic Büchi automaton recognizes the set of reals which are not of the form nb^p with $n \in \mathbb{N}$ and $p \in \mathbb{Z}$, since those reals are the ones whose encoding in base b contains an infinite number of non-0 digits. Furthermore, by [12], weak deterministic Büchi automata can be efficiently minimized.

We now recall some results about the above-mentioned logic. By [7], the logic $\text{FO}[\mathbb{R}; +, <, 1]$ admits quantifier elimination. By [17, Section 6], the set of reals which are $\text{FO}[\mathbb{R}; +, <, 1]$ -definable are the finite union of intervals with rational bounds. Those sets are called the *simple sets*.

Main results

It is shown that ideas given in [14] and [9] to create algorithms to decide properties of automata over integers can be adapted to decide properties of RVA. For examples, those ideas are used in Section 3 to give algorithms which decide whether a Büchi automaton recognizes a $\text{FO}[\mathbb{R}; +, <, 1]$ -definable set of tuple of reals, a $\text{FO}[\mathbb{R}, \mathbb{Z}; +, <]$ -definable set of tuple of real or a subsemigroup of $(\mathbb{R}^d, +)$ for some $d \in \mathbb{N}^{>0}$. However, those algorithms are inefficient.

It is then shown in Section 5 that it is decidable in linear time whether a RNA recognizes a $\text{FO}[\mathbb{R}; +, <, 1]$ -definable set, that is, a simple set. This algorithm does not return any false positive on weak deterministic Büchi automata which are not RNA. A false negative is also exhibited and it is explained why this case is more complicated than the case of RNA. A characterization of the minimal weak RNA which recognizes simple sets is also given.

Note that, if an automaton recognizes a simple set R , that is a finite union of intervals, the minimal number of intervals in the union is not polynomially bounded by the number of states of the automaton (this is shown in Example 2.1). It is shown in Section 6 that an existential (respectively, existential-universal) $\text{FO}[\mathbb{R}; +, <, 1]$ -formula which defines R is computable in quasi-quadratic (respectively quasi-linear) time.

1 Definitions

The definitions used in this paper are given in this section. Some basic lemmas are also given. Most definitions are standard.

1.1 Basic Notations

Let \mathbb{N} , \mathbb{Z} , \mathbb{Q} and \mathbb{R} denote the set of non-negative integers, integers, rationals and reals, respectively. For $R \subseteq \mathbb{R}$, let $R^{\geq 0}$ and $R^{>0}$ denote the set of non-negative and of positive elements of R , respectively. Let ω be the cardinality of \mathbb{N} . For $n \in \mathbb{N}$, let $[n]$ represent $\{0, \dots, n\}$. For $a, b \in \mathbb{R}$ with $a \leq b$, let $[a, b]$ denote the closed interval $\{r \in \mathbb{R} \mid a \leq r \leq b\}$, and let (a, b) denote the open interval $\{r \in \mathbb{R} \mid a < r < b\}$. Similarly, let $(a, b]$ (respectively, $[a, b)$) be the half-open interval equals to the union of (a, b) and of $\{b\}$ (respectively, $\{a\}$). For $r \in \mathbb{R}$ let $\lfloor r \rfloor$ be the greatest integer less than or equal to r .

1.2 Finite and infinite words

An alphabet is a finite set, its elements are called letters. A finite (respectively infinite) word of alphabet A is a finite (respectively infinite) sequence of letters of A . That is, a function from $[n]$ to A for some $n \in \mathbb{N}$ (respectively from \mathbb{N} to A). A set of finite (respectively infinite) word of alphabet A is called a language (respectively, an ω -language) of alphabet A . The empty word is denoted ϵ .

Let w be a word. Let $|w| \in \mathbb{N} \cup \{\omega\}$ denote the length of w . For v a finite word, let $u = vw$ be the concatenation of v and of w , that is, the word of length $|v| + |w|$ such that $u[i] = v[i]$ for $i < |v|$ and $u[|v| + i] = w[i]$ for $i < |w|$. For $n < |w|$, let $w[n]$ denote the n -th letter of w . Let $w[< n]$ denote the *prefix* of w of length n , that is, the word u of length n such that $w[i] = u[i]$ for all $i \in [n - 1]$. Similarly, let $w[\geq n]$ denote the *suffix* of w without its n -th first letters, that is, the word u such that $u[i] = w[i + n]$ for all $i \in [n - \omega]$. Note that $w = w[< i] w[\geq i]$ for all $i < |w|$.

Let L be a language of finite word and let L' be either an ω -languages or a language of finite words. Let LL' be the set of concatenations of the words of L and of L' . For $i \in \mathbb{N}$, let L^i be the concatenations of i words of L . Let $L^* = \bigcup_{i \in \mathbb{N}} L^i$ and $L^+ = \bigcup_{i \in \mathbb{N}^{>0}} L^i$. If L is a language which does not contains the empty word, let L^ω be the set of infinite sequences of elements of L .

1.2.1 Encoding of real numbers

Let us now consider the encoding of numbers in an integer base $b \geq 2$. Let Σ_b be equal to $[b - 1]$, it is the set of digits and let $\Sigma_b^* = \Sigma_b \cup \{\star\}$. The base b is fixed for the remaining of this paper. Two alphabets are considered in this paper: Σ_b and Σ_b^* .

Let $[\cdot]_b$ denote the function which sends a finite or infinite word of alphabet Σ_b^* to the integer or to the real it represents. Formally, for $w \in \Sigma_b^*$:

$$[w]_b^I = \sum_{i=0}^{|w|-1} b^{|w|-1-i} w[i].$$

For $w \in \Sigma_b^\omega$,

$$[w]_b^R = \sum_{i \in \mathbb{N}} b^{-i-1} w[i].$$

Let w be an ω -word with exactly one \star . It is of the form $w = w_I \star w_F$, with $w_I \in \Sigma_b^*$ and $w_F \in \Sigma_b^\omega$. The word w_I is called the natural part of w and the ω -word w_F is called its fractional part. Then :

$$[w_I \star w_F]_b^R = [w_I]_b^I + [w_F]_b^F.$$

Finally, $[w]_b^R$ is undefined if w contains at least two letters \star . There is no ambiguity in the definition of $[\cdot]_b^I$ since the four domains of definitions partition $(\Sigma_b^*)^\omega$. Note that $[w_I]_b^R \in \mathbb{N}$, $[w_F]_b^R \in [0, 1]$ and $[w]_b^F = [w_I]_b^R + [w_F]_b^R$. Examples of numbers with their base 2 encodings are now given.

Example 1.1.

$$\begin{array}{cccccc} [(10)^\omega]_2^R = \frac{2}{3} & [(01)^\omega]_2^R = \frac{1}{3} & [0(10)^\omega]_2^R = \frac{1}{3} & [0(1)^\omega]_2^R = \frac{1}{2} & [1(0)^\omega]_2^R = \frac{1}{2} \\ [10]_2^R = 2 & [1]_2^R = 1 & [01]_2^R = 1 & [\epsilon]_2^R = 0 & [00000]_2^R = 0 \\ [10 \star (10)^\omega]_2^R = \frac{8}{3} & [\star 0(1)^\omega]_2^R = \frac{1}{2} & & & [00000 \star 1(0)^\omega]_2^R = \frac{1}{2}. \end{array}$$

Some properties of concatenation and encodings of reals are now stated. The proof of the lemma is straightforward from the definition.

Lemma 1.2. *Note that for all $v \in \Sigma_b^*$, $w \in \Sigma_b^\omega$ and $a \in \Sigma_b$:*

$$\begin{array}{llll} [aw]_b^R &= \frac{a+[w]_b^R}{b}, & [av]_b^I &= ab^{|v|} + [v]_b^I, & [w]_b^R &= [0 \star w]_b^R, \\ [va]_b^I &= b[v]_b^I + a, & \text{and} & [av \star w]_b^R &= ab^{|v|} + [v \star w]_b^R. \end{array}$$

1.2.2 Encoding of rationals

In this section, some basic facts about rationals are recalled (see e.g. [8]). The rationals are exactly the numbers which admit encodings in base b of the form $u \star vv^\omega$ with $u, v \in \Sigma_b^*$ and $w \in \Sigma_b^+$. Rationals of the form nb^p , with $n \in \mathbb{N}$ and $p \in \mathbb{Z}$, admit exactly two encodings in base b without leading 0 in the natural part. If $p < 0$, the two encodings are of the form $u \star va(b-1)^\omega$ and $u \star v(a+1)0^\omega$, with $u, v \in \Sigma_b^*$ and $a \in [b-2]$. Otherwise, if $p \geq 0$, the two encodings are of the form $ua(b-1)^q \star (b-1)^\omega$ and $u(a+1)0^q \star 0^\omega$ with $u \in \Sigma_b^*$, $a \in [b-2]$ and $q \in \mathbb{N}$. The rationals which are not of the form nb^p admit exactly one encoding in base b without leading 0 in the natural part.

1.2.3 Encoding of sets of reals

In this section, relations between languages and set of reals are recalled.

Given a language L in Σ_b^ω or in $\Sigma_b^* \star \Sigma_b^\omega$, let $[L]_b^R$ be the set of reals admitting an encoding in base b in L . The language L is said to be an encoding in base b of the set of reals $[L]_b^R$. Reciprocally, given a set $R \subseteq \mathbb{R}^{\geq 0}$ of reals, $L_b(R)$ is the set of all encodings in base b of the elements of R .

Following [11], a language L is said to be *saturated* if for any number r which admits an encoding in base b in L , all encoding in base b of r belongs to L . The saturated languages are of the form $L_b(R)$ for $R \subseteq \mathbb{R}^{\geq 0}$. Note that

$[L_b(R)]_b^R = R$ for all sets $R \subseteq \mathbb{R}^{\geq 0}$. Note also that $L \subseteq L_b([L]_b^R)$, and the subset relation is an equality if and only if L is saturated. In general, a set of reals may have infinitely many encodings in base b . For example, for $I \subseteq \mathbb{N}$ an arbitrary set, $\{0, 1\}^\omega \setminus \{0^i 1^\omega \mid i \in I\}$ is an encoding in base 2 of the language of the simple set $[0, 1]$. An example of set of reals is now given.

Example 1.3. Let $L = \Sigma_2^* 0^\omega$ and $L' = \Sigma_2^* 0 \Sigma_2^* (0^\omega + 1^\omega)$. Both $[L]_2^R$ and $[L']_2^R$ are $R = \{\frac{n}{2^p} \mid n, p \in \mathbb{N}, n < 2^p\}$, but only L' is saturated. Therefore $L_2(R) = L'$.

1.3 Deterministic Büchi automata

This paper deals with Deterministic Büchi automata. This notion is now defined.

A *Deterministic Büchi automaton* is a 5-tuple $\mathcal{A} = (Q, A, \delta, q_0, F)$, where Q is a finite set of states, A is an alphabet, $\delta \subseteq Q \times A \times Q$ is the transition relation, $q_0 \in Q$ is the initial states and $F \subseteq Q$ is the set of accepting states. A state belonging to $Q \setminus F$ is said to be a *rejecting state*.

An example of deterministic Büchi automaton is now given. This example is used thorough this paper to illustrate properties of Büchi automaton reading set of real numbers.

Example 1.4. Let $R = (\frac{1}{3}, 2] \cup (\frac{8}{3}, 3] \cup (\frac{11}{3}, \infty]$. The set of encodings in base 2 of reals of R is recognized by the automaton pictured in Figure 1.

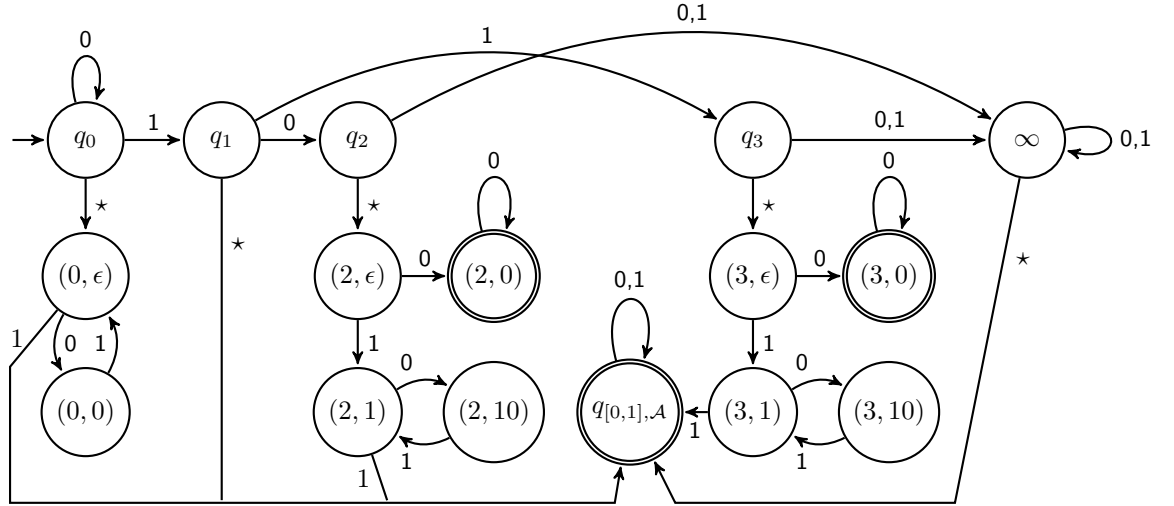


Figure 1: Automaton \mathcal{A}_R of Example 1.4

From now on in this paper, all automata are assumed to be deterministic. The function δ is implicitly extended on $Q \times A^*$ by $\delta(q, \epsilon) = q$ and $\delta(q, aw) = \delta(\delta(q, a), w)$ for $a \in A$ and $w \in A^*$.

Let \mathcal{A} be an automaton and w be an infinite word. A *run* π of \mathcal{A} on w is a mapping $\pi : \mathbb{N} \mapsto Q$ such that $\pi(0) = q_0$ and $\delta(\pi(i), w[i]) = \pi(i+1)$ for all $i < |w|$. Let $\text{inf}(\pi)$ be the set of states of Q that occur infinitely often in the run π . A run π on an ω -word is said to be *accepting* if $\text{inf}(\pi) \cap F \neq \emptyset$. Equivalently, the run is accepting if there exists a state $q \in F$ such that there is an infinite number of $i \in \mathbb{N}$ such that $\pi(i) = q$. Example 1.4 is now resumed.

Example 1.5. Let \mathcal{A} be the automaton pictured in Figure 1. The run of \mathcal{A} on $011 \star (10)^\omega$ is

$$(q_0, q_0, q_1, q_3, (3, \epsilon), (3, 1), (3, 10), \dots)$$

with the two last states repeated infinitely often. The Büchi automaton \mathcal{A} does not accept $011 \star (10)^\omega$ since this run does not contain any accepting state.

The run of \mathcal{A} on $\star 1^\omega$ is $(q_0, (0, \epsilon), q_{[0,1], \mathcal{A}}, \dots)$ with the last state repeated infinitely often. The Büchi automaton \mathcal{A} accepts $\star 1^\omega$ since the accepting state $q_{[0,1], \mathcal{A}}$ appears infinitely often in the run.

Let \mathcal{A} be a finite automaton. Let $L_\omega(\mathcal{A})$ be the set of infinite words w such that a run of \mathcal{A} on w is accepting. An ω -language is said to be *recognizable* if it is recognized by a Büchi automaton. Example 1.4 is now resumed.

Example 1.6. Let \mathcal{A} be the Büchi automaton pictured in Figure 1. It recognizes the language of encodings in base 2 of the reals of $(\frac{1}{3}, 2] \cup (\frac{8}{3}, 3] \cup (\frac{11}{3}, \infty]$. It is explained in Example 5.3 how this automaton was computed.

For $q \in Q$, let \mathcal{A}_q be (Q_q, A, δ, q, F_q) , where Q_q is the set of states of Q accessible from q , and $F_q = F \cap Q_q$.

Note that, if there are no finite word w such that $\delta(q_0, w) = q_0$, then $Q_q \subsetneq Q$ for all $q \neq q_0$. Note also that, if $w \in A^*$ is such that $\delta(q_0, w) = q$ then a word $w' \in A^\omega$ is accepted by \mathcal{A}_q if and only if ww' is accepted by \mathcal{A} .

1.3.1 Accessibility and recurrent states

When the notions of initial and of accepting states are ignored, an automaton can be considered as a directed labelled graph. Some definitions related to this graph are introduced in this section.

A state q is said to be *accessible* from a state q' if there exists a finite non-empty word w such that $\delta(q', w) = q$. Following [12], a state q is said to be *recurrent* if it is accessible from itself and *transient* otherwise. Transient states are called *trivial* in [?]. The *strongly connected component* of a recurrent state q is the set of states q' such that q' is accessible from q and q is accessible from q' . A strongly connected component C is said to be a *leaf* if for all $a \in A$, for all $q \in C$, $\delta(q, a) \in C$. Let C be a strongly connected component. It is said to be a *cycle* if for each $q \in C$, there exists a unique $s_q \in A$ such that $\delta(q, s_q) \in C$. Example 1.4 is now resumed.

Example 1.7. The transient states of the automaton pictured in Figure 2 are $q_0, q_1, q_2, q_3, (2, \epsilon)$ and $(3, \epsilon)$. All other states are recurrent. The cycles are $\{q_0\}$, $\{(0, \epsilon), (0, 0)\}$, $\{(2, 0)\}$, $\{(2, 1), (2, 10)\}$, $\{(3, 0)\}$ and $\{(3, 1), (3, 10)\}$. The strongly connected component which are not cycles are $q_{\emptyset, \mathcal{A}}, q_{\infty, \mathcal{A}}$ and $q_{[0,1], \mathcal{A}}$.

The following lemma allows to consider recurrent states in any run which is long-enough.

Lemma 1.8. *Let \mathcal{A} be a Büchi automaton with n states let w be an ω -word and let π be the run of \mathcal{A} on w . Let $N \subseteq \mathbb{N}$ be a set of cardinal at least $(n+1)$. Then there is $i < i'$ belonging to N such that $\pi(i) = \pi(i')$ is a recurrent state.*

Proof. Since the cardinality of N is greater than the number of state, by the pigeonhole principle, there exists $i < i'$ belonging to N such that $\pi(i) = \pi(i')$. Let w' be the factor of w containing the letters $i+1$ to i' , then $\delta(\pi(i), w') = \pi(i') = \pi(i)$, therefore, the state $\pi(i)$ is recurrent, with i belonging to N . \square

1.3.2 Quotients, Morphisms and Weak Büchi Automata

In this section, the notion of quotient of automata and of morphism of automata are introduced. A class of automata admitting minimal quotient is then introduced.

Definition 1.9 (Morphism of Büchi automata, Quotient). Let $\mathcal{A} = (Q, A, \delta, q_0, F)$ and $\mathcal{A}' = (Q', A, \delta', q'_0, F')$ be two Büchi automata over the same alphabet. A surjective function $\mu : Q \rightarrow Q'$ is a *morphism* of Büchi automata if and only if:

1. $\mu(q_0) = q'_0$,
2. for each $q \in Q$, $L_\omega(\mathcal{A}_q) \neq L_\omega(\mathcal{A}_{q'})$.

The Büchi automaton \mathcal{A}' is said to be a *quotient* of \mathcal{A} if there exists a morphism from \mathcal{A} to \mathcal{A}' .

The notion of minimal Büchi automaton is now introduced.

Definition 1.10 (Minimal Büchi automaton). Let $\mathcal{A} = (Q, \Sigma_b^*, \delta, q_0, F)$ be a Büchi automaton. It is said to be minimal if for each distinct states $q, q' \in Q$, $L_\omega(\mathcal{A}_q) = L_\omega(\mathcal{A}_{q'})$.

In general, Büchi automata does not admit minimal quotient. A class of Büchi automata admitting minimal quotient is now introduced.

Definition 1.11 (Weak automata). Let $\mathcal{A} = (Q, \Sigma_b^*, \delta, q_0, F)$ be a Büchi automaton. It is said to be *weak* if for each recurrent accepting state q of \mathcal{A} , all states of the strongly connected components of q are accepting.

An ω -language is said to be *weakly recognizable* if it is recognized by a weak Büchi automaton.

The main theorem concerning quotient of weak Büchi automata is now recalled.

Theorem 1.12 ([12]). *Let \mathcal{A} be a weak Büchi automaton with n states such that all states of \mathcal{A} are accessible from its initial state. Let c be the cardinality of A . There exists a minimal weak Büchi automaton \mathcal{A}' such that there exists a morphism of automaton μ from \mathcal{A} to \mathcal{A}' . The automaton \mathcal{A}' and the morphism μ are computable in time $O(n \log(n)c)$ and space $O(nc)$.*

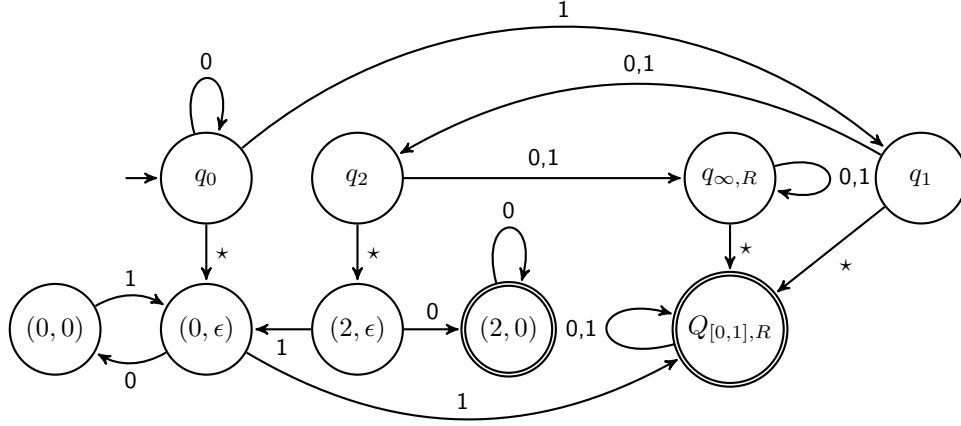


Figure 2: Minimal quotient of automaton \mathcal{A}_R of Figure 1

It follows easily from Property (??) that, for all $w \in A^*$, $\delta'(\mu(q), w) = \mu(\delta(q, w))$. Example 1.4 is now resumed.

Example 1.13. Let \mathcal{A}_R be the automaton pictured in Figure 1. Its minimal quotient is pictured in Figure 2.

The following lemma shows that each strongly connected component of a quotient by a morphism μ from an automaton \mathcal{A} is the image of a strongly connected component of \mathcal{A} .

Lemma 1.14. *Let $\mathcal{A} = (Q, A, \delta, q_0, F)$ and $\mathcal{A}' = (Q', \Sigma_b, \delta', q'_0, F')$ be two Büchi automata. Let μ be a morphism from \mathcal{A} to \mathcal{A}' . Let C' be a strongly connected component of \mathcal{A}' . There exists a strongly connected component $C \subseteq Q$ such that $\mu(C) = C'$ and such that, for all $q \in Q \setminus C$ accessible from C , $\mu(q) \notin C'$.*

In order to prove this lemma, two other lemmas are required.

Lemma 1.15. *Let \mathcal{A} , \mathcal{A}' , C' and μ as in Lemma 1.14. Let C be a strongly connected component of \mathcal{A} . Either $\mu(C) \cap C' = \emptyset$ or $\mu(C) \subseteq C'$.*

Proof. Let us assume that $\mu(C) \cap C' \neq \emptyset$ and let us prove that $\mu(C) \subseteq C'$. That is, let $q \in C$ and let us prove that $\mu(q) \in C'$.

Since $\mu(C) \cap C' \neq \emptyset$, there exists $q' \in \mu(C) \cap C'$. Since $q' \in \mu(C)$, there exists $p \in C$ such that $\mu(p) = q'$. Since p and q belong to the same strongly connected component, there exists two non-empty finite words v and w such that $\delta(p, v) = q$ and $\delta(q, w) = p$. Therefore $\delta'(\mu(p), v) = \mu(\delta(p, v)) = \mu(q)$ and $\delta'(\mu(q), w) = \mu(\delta(q, w)) = \mu(q)$. Therefore $\mu(q')$ is accessible from $\mu(q)$ and $\mu(q')$ is accessible from $\mu(q)$. Hence $\mu(q)$ belongs to the strongly connected component of p . That is, $\mu(q)$ belongs to C' . \square

Lemma 1.16. *Let \mathcal{A} , \mathcal{A}' , C' and μ as in Lemma 1.14. Let $q \in Q$ such that $\mu(q) \in C'$. There exists a strongly connected component C of \mathcal{A} , accessible from q , such that $\mu(C) \subseteq C'$.*

Proof. Since $\mu(q) \in C'$, the state $\mu(q)$ is recurrent, therefore there exists a non-empty word w such that $\delta'(\mu(q), w) = \mu(q)$. Let us prove by induction on $i \in \mathbb{N}$ that $\mu(\delta(q, w^i)) = \mu(q)$. The case $i = 0$ is trivial, let us assume that the hypothesis holds for $i \in \mathbb{N}$ and let us prove that the induction hypothesis holds for $i + 1$. It suffices to see that

$$\mu(\delta(q, w^{i+1})) = \delta'(\mu(q), w^{i+1}) = \delta'(\delta'(\mu(q), w^i), w) = \delta'(\mu(q), w) = \mu(q)$$

By Lemma 1.8, there exists $i \in \mathbb{N}$ such that $\delta(q, w^i)$ is recurrent. Let C be the strongly connected of $\delta(q, w^i)$. Since $\delta(q, w^i) \in C$ and $\mu(\delta(q, w^i)) \in C'$, by Lemma 1.15, it implies that $\mu(C) \subseteq C'$. Since C is accessible from q and $\mu(C) \subseteq C'$, the lemma is satisfied. \square

Lemma 1.14 is now proven.

Proof of Lemma 1.14. Let $\mu^{-1}(C') \subseteq Q$ be the set of states q such that $\mu(q) \in C'$. By definition of morphism, μ is surjective, hence $\mu^{-1}(C')$ is not empty. Let q be a state belonging to $\mu^{-1}(C')$. By Lemma 1.16, it implies that there exists a strongly connected component $C \subseteq \mu^{-1}(C')$. Since $\mu^{-1}(C')$ is finite, there exists a strongly connected component C such that no other strongly connected component of $\mu^{-1}(C')$ is accessible from C . By Lemma 1.16 it implies that no state of $\mu^{-1}(C') \setminus C$ is accessible from C .

Let us prove that $\mu(C) = C'$. Since, by hypothesis $\mu(C) \subseteq C'$, it remains to prove that $\mu(C) \supseteq C'$. Let $q' \in C'$ and let us prove that $q' \in \mu(C)$. Let $q \in C$. By hypothesis, $\mu(C) \subseteq C'$, therefore $\mu(q) \in C'$. Since $\mu(q)$ and q' belong to the same strongly connected component C' , there exists a finite word w such that $\delta'(\mu(q), w) = q'$. Then $\mu(\delta(q, w)) = \delta'(\mu(q), w) = q'$. Since $\mu(\delta(q, w)) = q' \in C'$, $\delta(q, w) \in \mu^{-1}(C')$. Since $\delta(q, w) \in \mu^{-1}(C')$ and $\delta(q, w)$ is accessible from $q \in C$, by hypothesis on C , $\delta(q, w) \in C$. Therefore $q' = \mu(\delta(q, w)) \in \mu(C)$. \square

1.4 Logic

The logic $\text{FO}[\mathbb{R}; +, <, 1]$ used in this paper is introduced in this section. Note that, in order to avoid ambiguity between the mathematical equality and the formal equality of the logic, the symbol \doteq is used in first-order formulas.

Intuitively, FO stands for first-order. The first parameter \mathbb{R} means that the (free or quantified) variables are interpreted by real numbers. The $+$ and $<$ symbols mean that the function addition and the binary order relation over reals can be used in formulas. Finally, the last term, 1, means that the only constant is 1. The logic $\text{FO}[\mathbb{R}; +, <, 1]$ is denoted by \mathcal{L} in [7], where it is proved that this logic admits quantifier elimination. In this paper, most results deal with the quantifier-free, the existential fragment and the existential-universal fragment of $\text{FO}[\mathbb{R}; +, <, 1]$ denoted by $\Sigma_0[\mathbb{R}; +, <, 1]$, $\Sigma_1[\mathbb{R}; +, <, 1]$ and $\Sigma_2[\mathbb{R}; +, <, 1]$ respectively.

In the remaining of the paper, rationals are also used in the formulas. It does not change the expressivity, as all rational constants are $\Sigma_0[\mathbb{R}; +, <, 1]$ -definable. Let $\phi \in \text{FO}[\mathbb{R}; +, <, 1]$. The length of ϕ , denoted by $|\phi|$, is recursively defined as follows:

- The lengths of the constant $\frac{p}{q}$ is $\log p + 1 + \log q$.
- The length of a sum $t_1 + t_2$ is $1 + |t_1| + |t_2|$.
- The length of a multiplication by a rational constant $\frac{p}{q}t$ is $\left\lceil \frac{p}{q} \right\rceil + |t|$.
- The length of an (in)equality is the sum of the length of the terms on both side, plus one, that is $|t_1 < t_2| = |t_1 \dot{=} t_2| = 1 + |t_1| + |t_2|$.
- The length of Boolean combination and of quantification are 1 plus the length of its subterms, that is $|\phi \vee \psi| = |\phi \wedge \psi| = 1 + |\phi| + |\psi|$ and $|\exists x.\phi| = |\forall x.\phi| = |\neg\phi| = 1 + |\phi|$.

1.4.1 First-order definable sets of reals

In this section, notations are introduced for the kind of sets studied in this paper: the $\text{FO}[\mathbb{R}; +, <, 1]$ -definable sets.

Following [17, Section 6], the $\text{FO}[\mathbb{R}; +, <, 1]$ -definable sets are called the *simple sets*. By [17, Section 6], those sets are the finite union of intervals with rational bounds. It implies that there exists an integer t_R such that for all $x, y \geq t_R$, x belongs to R if and only if y belongs to R . The least such integer t_R is called the *threshold of R* .

Note that every closed and half-closed intervals is the union of an open interval and of singletons, hence it can be assumed that any simple set R is of the form

$$R = \bigcup_{i=0}^{I-1} (\rho_{i,\mathfrak{L}}, \rho_{i,\mathfrak{R}}) \cup \bigcup_{i=0}^{J-1} \{\rho_{i,\mathfrak{S}}\},$$

with $\rho_{i,\mathfrak{L}}, \rho_{i,\mathfrak{S}} \in \mathbb{Q}^{\geq 0}$ and $\rho_{i,\mathfrak{R}} \in \mathbb{Q}^{\geq 0} \cup \{\infty\}$. The $\rho_{i,\mathfrak{L}}$ are the left bound, the $\rho_{i,\mathfrak{R}}$ are the right bound and the $\rho_{i,\mathfrak{S}}$ are the singletons. Without loss of generality, it is assumed that the intervals are disjoint and in increasing order.

Example 1.4 is now resumed.

Example 1.17. Let $R = (\frac{1}{3}, 2] \cup (\frac{8}{3}, 3] \cup (\frac{11}{3}, \infty]$ as in Example 1.4. Then t_R is 4, $R_0 = (\frac{1}{3}, 1]$, $R_1 = [0, 1]$, and $R_2 = R_3 = \{0\} \cup [\frac{2}{3}, 1]$. Furthermore, $I = 3$, $J = 2$, $\rho_{1,\mathfrak{L}} = \frac{1}{3}$, $\rho_{2,\mathfrak{R}} = 2$, $\rho_{2,\mathfrak{L}} = \frac{8}{3}$, $\rho_{2,\mathfrak{R}} = 3$, $\rho_{3,\mathfrak{L}} = \frac{11}{3}$, $\rho_{3,\mathfrak{R}} = \infty$, $\rho_{1,\mathfrak{S}} = 2$ and $\rho_{2,\mathfrak{S}} = 3$.

2 Automata reading reals

Automata recognizing encoding of set of reals are considered in this section. The notion of Real Number Automata and of Fractional Number Automata are

introduced in Section 2.1. Some sets of states of the automata reading encoding of set of reals are considered in Section 2.2.

2.1 Real and Fractional Number Automata

In this section, the automata reading saturated languages are considered.

Following [2], a Büchi automaton of alphabet \star is said to be a Real Number Automaton (RNA) if

- all words accepted by \mathcal{A} contains exactly one \star , and
- the language $L_\omega(\mathcal{A})$ is saturated.

The Büchi automata pictured in 1 and 2 are RNA. Clearly, the RNAs are the Büchi automata which recognizes saturated languages of $\Sigma_b^* \star \Sigma_b^\omega$. Similarly, the name of Fractional Number Automata (FNA) is given to the Büchi automata of alphabet Σ_b recognizing a saturated language.

A weak Büchi automaton which is a RNA or a FNA is said to be a *weak RNA* or a *weak FNA* respectively. An example of FNA is now given. This example shows that the number of intervals required to describe a set is not polynomially bounded by the number of states of the automaton recognizing this set.

Example 2.1. For every non-negative integer n , let R_n be $\{m2^{-n} \mid m \in [2^n]\}$. It is the set of reals which admit an encoding w in base b whose suffixes $w[\geq n]$ are either equal to 0^ω or to 1^ω . This set can not be described with less than 2^{n-1} intervals and is recognized by the automaton \mathcal{A}_n with $n+3$ states:

$$\mathcal{A}_n = (\{q_i \mid i \in [n]\} \cup \{q_{n+1,0}, q_{n+1,1}, q_{\emptyset, \mathcal{A}}\}, \Sigma_b, \delta, q_0, \{q_{n+1,0}, q_{n+1,1}\}),$$

where the transition function is such that, for $a \in \Sigma_2$:

$$\begin{aligned} \delta(q_i, a) &= q_{i+1} \text{ for } i \in [n-1], & \delta(q_n, a) &= q_{n+1,a}, \\ \delta(q_{n+1,a}, a) &= q_{n+1,a} & \delta(q_{n+1,a}, 1-a) &= q_{\emptyset, \mathcal{A}} \text{ and } \delta(q_{\emptyset, \mathcal{A}}, a) = q_{\emptyset, \mathcal{A}}. \end{aligned}$$

The automaton \mathcal{A}_3 is pictured in Figure 3, without the state $q_{\emptyset, \mathcal{A}}$.

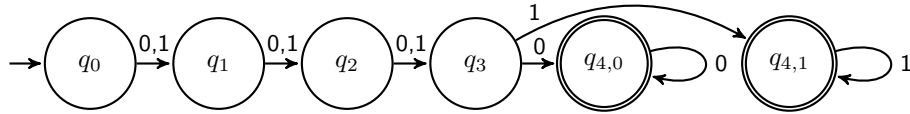


Figure 3: The automaton \mathcal{A}_3 of Example 2.1, accepting $\{0 = \frac{0}{8}, \frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \frac{4}{8}, \frac{5}{8}, \frac{6}{8}, \frac{7}{8}, \frac{8}{8} = 1\}$

For \mathcal{A} a Büchi automaton of alphabet Σ_b (respectively, Σ_b^*). Let $[\mathcal{A}]_b^{\mathbb{R}} = [L_\omega(\mathcal{A})]_b^{\mathbb{R}}$. It is a subset of $[0, 1]$ (respectively, $\mathbb{R}^{\geq 0}$). It is said that \mathcal{A} recognizes $[\mathcal{A}]_b^{\mathbb{R}}$.

It should be noted that two distinct minimal weak Büchi automata may recognize the same set of reals. Indeed, they may recognize two distinct languages which are two encodings of the same set of reals. At least one of those languages is not saturated. Note however that two distinct minimal RNA or RFA accept distinct sets of reals.

2.2 Some sets of states of RNA and of FNA

Five sets of states of Büchi automata are used through this paper. Those sets are introduced and studied in this section.

Definition 2.2 ($Q_{\emptyset, \mathcal{A}}, Q_{[0,1], \mathcal{A}}, Q_{\infty, \mathcal{A}}, Q_{I, \mathcal{A}}$ and $Q_{F, \mathcal{A}}$). Let \mathcal{A} be an automaton over alphabet Σ_b^* or Σ_b .

- Let $Q_{\emptyset, \mathcal{A}}$ be the set of states q such that \mathcal{A}_q recognizes the empty language.
- Let $Q_{[0,1], \mathcal{A}}$ be the set of states q such that \mathcal{A}_q recognizes $\Sigma_b^\omega = L_b([0, 1])$.
- Let $Q_{\infty, \mathcal{A}}$ be the set of states q such that \mathcal{A}_q recognizes the language $\Sigma_b^* \star \Sigma_b^\omega = L_b([0, \infty))$.
- Let $Q_{I, \mathcal{A}}$ be the set of states q such that \mathcal{A}_q recognizes a subset of $\Sigma_b^* \star \Sigma_b^\omega$.
- Let $Q_{F, \mathcal{A}}$ be the set of states q such that \mathcal{A}_q recognizes a subset of Σ_b^ω .

Example 1.13 is now resumed.

Example 2.3. Let \mathcal{A} be the automaton pictured in Figure 2. Let $q_{\emptyset, \mathcal{A}}$ be the state $\delta((2, 0), 1)$, which is not pictured in Figure 2. Then $Q_{[0,1], \mathcal{A}} = \{q_{[0,1], R}\}$, $Q_{\infty, \mathcal{A}} = \{q_{\infty, R}\}$ and $Q_{\emptyset, \mathcal{A}} = \{q_{\emptyset, \mathcal{A}}\}$. Furthermore, $Q_{I, \mathcal{A}} = \{q_0, q_1, q_2, q_{\infty, R}, q_{\emptyset, R}\}$, its elements are represented in the top row, of Figure 2. Finally, $Q_{F, \mathcal{A}} = \{(2, \epsilon), (2, 0), (0, \epsilon), (0, 0), q_{[0,1], R}, q_{\emptyset, R}\}$. Its elements are pictured in the second row of Figure 2.

The following lemma is straightforward from the definition.

Lemma 2.4. *In a minimal weak Büchi automaton, the sets $Q_{\emptyset, \mathcal{A}}$, $Q_{[0,1], \mathcal{A}}$ and $Q_{\infty, \mathcal{A}}$ are either singletons or the empty set.*

In a minimal weak Büchi automaton \mathcal{A} , let $q_{\emptyset, \mathcal{A}}$, $q_{[0,1], \mathcal{A}}$ and $q_{\infty, \mathcal{A}}$ denote the only state q such that \mathcal{A}_q recognizes the languages \emptyset , Σ_b^ω and $\Sigma_b^* \star \Sigma_b^\omega$ respectively. In an automaton of alphabet Σ_b , all states belong to $Q_{F, \mathcal{A}}$.

The following lemma states that those five sets are linear time computable.

Lemma 2.5. *Let \mathcal{A} be a Büchi automaton with n states. Then the sets $Q_{\emptyset, \mathcal{A}}$, $Q_{I, \mathcal{A}}$ and $Q_{F, \mathcal{A}}$ are computable in time $O(nb)$. If \mathcal{A} is weak, the sets $Q_{[0,1], \mathcal{A}}$ and $Q_{\infty, \mathcal{A}}$ are computable in time $O(nb)$.*

Proof. Tarjan's algorithm [16] can be used to compute the set of strongly connected component in time $O(nb)$, and therefore the set of recurrent states. Furthermore, it is easy to associate in linear time to each state its set of predecessors. Let p_q be the number of predecessors of a state q .

Let us first explain how to compute the set $Q_{\emptyset, \mathcal{A}}$. Note that $Q \setminus Q_{\emptyset, \mathcal{A}}$ is the set of states q such that \mathcal{A}_q accepts some ω -word. Hence $Q \setminus Q_{\emptyset, \mathcal{A}}$ is the smallest set containing all accepting recurrent states and is closed under taking predecessors. Therefore $Q_{\emptyset, \mathcal{A}}$ is the greatest set which does not contain the accepting recurrent states and is closed under taking successor. It can thus be computed by a fixed-point algorithm. The algorithm is now given.

Two sets S and S' are used by the algorithm. The set S represents $Q_{\emptyset, \mathcal{A}}$. The set S' is the set of states of $Q_{\emptyset, \mathcal{A}}$ which must be processed by the fixed-point algorithm. The algorithm initializes the set S to Q and initializes the set S' to the empty set. The algorithm runs on each recurrent state q . For each state q , if q is accepting, then q is removed from S and added to S' . The algorithm then runs on each element q of S' . For each state q , the algorithm removes q from S' and runs on each predecessors q' of q . For each q' , if q' is in S , then q' is removed from S and added to S' . Finally, when S' is empty, the algorithm halts and $Q_{\emptyset, \mathcal{A}}$ is the value of S .

Let us now consider the computation time of this algorithm. At most n states are added to S' , and each state is added at most once. For each state q added to S' , each of its c_q predecessor is considered in constant time. Thus the algorithm runs in time $O\left(n + \sum_{q \in Q} c_q\right) = O(nb)$.

It is now explained how to compute $Q_{F, \mathcal{A}}$ and $Q_{I, \mathcal{A}}$. Let Q_0 , Q_1 and Q_2 be the set of states accepting a words with at least 0, 1 and 2 \star 's respectively. Then $Q_{F, \mathcal{A}}$ is equal to $(Q_0 \setminus Q_1) \cup Q_{\emptyset, \mathcal{A}}$ and $Q_{I, \mathcal{A}}$ is equal to $(Q_1 \setminus Q_2) \cup Q_{\emptyset, \mathcal{A}}$. Let us now explain how to compute the sets Q_0 , Q_1 and Q_2 . The set Q_0 is the smallest set containing all accepting recurrent states and closed under taking predecessors. The state Q_1 is the smallest set containing the predecessors of Q_0 by the letter \star and closed under taking predecessors. Similarly, the set Q_2 is the smallest set containing the predecessors of Q_1 by the letter \star and closed under taking predecessors. Those three sets are computable by a fixpoint algorithm similar to the one computing $Q_{\emptyset, \mathcal{A}}$. It is thus computable in time $O(nb)$.

Let us now assume that the Büchi automaton \mathcal{A} is weak. It is now explained how to compute $Q_{[0,1], \mathcal{A}}$ and $Q_{\infty, \mathcal{A}}$. Note that $Q_{[0,1], \mathcal{A}} \subseteq Q_{F, \mathcal{A}}$ and that $Q_{F, \mathcal{A}} \setminus Q_{[0,1], \mathcal{A}}$ is the set of states $q \in Q_{F, \mathcal{A}}$ such that there exists an ω -word $w \in \Sigma_b^\omega$ which is not accepted by \mathcal{A}_q . Therefore, $Q_{F, \mathcal{A}} \setminus Q_{[0,1], \mathcal{A}}$ is the smallest subset of $Q_{F, \mathcal{A}}$ containing non-accepting recurrent state and closed under taking predecessors by Σ_b . Similarly, note that $Q_{\infty, \mathcal{A}} \subseteq Q_{I, \mathcal{A}}$ and that $Q_{I, \mathcal{A}} \setminus Q_{\infty, \mathcal{A}}$ is the set of states q such that there is an infinite word of the form $\Sigma_b^* \star \Sigma_b^\omega$ which is not accepted by \mathcal{A}_q . Therefore $Q_{I, \mathcal{A}} \setminus Q_{\infty, \mathcal{A}}$ is the smallest subset of $Q_{I, \mathcal{A}}$ containing the predecessors of $Q_{F, \mathcal{A}} \setminus Q_{[0,1], \mathcal{A}}$ by \star and closed under taking

predecessors by Σ_b . The computation of $Q_{I,\mathcal{A}}$ and of $Q_{\infty,\mathcal{A}}$ is thus similar to the computation of $Q_{\emptyset,\mathcal{A}}$. \square

This lemma admits the following corollary.

Corollary 2.6. *It is decidable in time $O(nb)$ whether a Büchi automaton with n states recognizes a subset of $\Sigma_b^* \star \Sigma_b^\omega$ or of Σ_b^ω .*

Proof. By definition of $Q_{I,\mathcal{A}}$ (respectively $Q_{F,\mathcal{A}}$), the automaton recognizes a subset of $\Sigma_b^* \star \Sigma_b^\omega$ (respectively of Σ_b^ω) if and only if its initial state belongs to $Q_{I,\mathcal{A}}$ (respectively $Q_{F,\mathcal{A}}$). By Lemma 2.5, it is testable in time $O(bn)$. \square

The following lemma gives a relation between the set of states introduced in Example 2.2 and morphisms of automata.

Lemma 2.7. *Let $\mathcal{A} = (Q, A, \delta, q_0, F)$ and $\mathcal{A}' = (Q', \Sigma_b, \delta', q'_0, F')$ be two Büchi automata. Let $\mu : Q \rightarrow Q'$ be a morphism of Büchi automaton. Then $\mu(Q_{\emptyset,\mathcal{A}}) = Q_{\emptyset,\mathcal{A}'}$, $\mu(Q_{F,\mathcal{A}}) = Q_{F,\mathcal{A}'}$, $\mu(Q_{\infty,\mathcal{A}}) = Q_{\infty,\mathcal{A}'}$, $\mu(Q_{I,\mathcal{A}}) = Q_{I,\mathcal{A}'}$, and $\mu(Q_{[0,1],\mathcal{A}}) = Q_{[0,1],\mathcal{A}'}$.*

Proof. The proof is done for the first equality: $\mu(Q_{\emptyset,\mathcal{A}}) = Q_{\emptyset,\mathcal{A}'}$. All other cases are similar. Let $q' \in Q'$, and let us prove that $q' \in \mu(Q_{\emptyset,\mathcal{A}}) \iff q' \in Q_{\emptyset,\mathcal{A}'}$. It suffices to see that:

$$\begin{aligned}
q' \in \mu(Q_{\emptyset,\mathcal{A}}) &\iff \exists q \in Q. q \in Q_{\emptyset,\mathcal{A}} \quad \wedge \mu(q) = q' \\
&\iff \exists q \in Q. L_\omega(\mathcal{A}_q) = \emptyset \quad \wedge \mu(q) = q' \\
&\iff \exists q \in Q. L_\omega(\mathcal{A}'_{\mu(q)}) = \emptyset \wedge \mu(q) = q' \\
&\iff \exists q \in Q. L_\omega(\mathcal{A}'_{q'}) = \emptyset \quad \wedge \mu(q) = q' \\
&\iff L_\omega(\mathcal{A}'_{q'}) = \emptyset \\
&\iff q' \in Q_{\emptyset,\mathcal{A}'}.
\end{aligned}$$

\square

The following lemma gives a relation between the set of states introduced in Example 2.2 and transitions. All of the results follow easily from the definition of those sets.

Lemma 2.8. *Let $\mathcal{A} = (Q, \Sigma_b^*, \delta, q_0, F)$. Let $q \in Q$ and $a \in \Sigma_b^*$. Then $\delta(q, a)$ belongs to the sets indicated in Example 1.*

If q belongs to	then $\delta(q, a)$, for $a \in \Sigma_b$, belongs to	and $\delta(q, \star)$ belongs to:
$Q_{\emptyset,\mathcal{A}}$	$Q_{\emptyset,\mathcal{A}}$	$Q_{\emptyset,\mathcal{A}}$
$Q_{[0,1],\mathcal{A}}$	$Q_{[0,1],\mathcal{A}}$	$Q_{\emptyset,\mathcal{A}}$
$Q_{\infty,\mathcal{A}}$	$Q_{\infty,\mathcal{A}}$	$Q_{[0,1],\mathcal{A}}$
$Q_{I,\mathcal{A}}$	$Q_{I,\mathcal{A}}$	$Q_{F,\mathcal{A}}$
$Q_{F,\mathcal{A}}$	$Q_{F,\mathcal{A}}$	$Q_{\emptyset,\mathcal{A}}$

Table 1: Set of states and transitions

3 Three methods to prove decidability of automata problems

In this section, three methods are given. Those methods allow to prove that some problems over automata are decidable.

The method given in Section 3.1 is based on an algorithm of [9], which decide whether an integer automaton recognizes an ultimately periodic set of integer. The method given in Section 3.2 is based on an algorithm of [14], which decide whether an automaton reading tuples of integers recognizes a set of tuples of integers which is $\text{FO}[\mathbb{N}; +]$ -definable. Finally, the method given in Section 3.3 is based on an algorithm of [13], which decides in linear time whether a minimal automaton recognizes an ultimately periodic set of integer.

Those algorithms are easily adapted to other decision problem on automata reading vectors of integers or of reals. While the methods given in the first two sections are less efficient than the method of the last section, it seems interesting to give those method as they are very general and, as far as the author know, has not yet been given in full generality. For example, it is shown that those methods allow to prove that it is decidable:

- whether an automaton recognizes a set of real which is $\text{FO}[\mathbb{R}; +, <, 1]$ -definable or $\text{FO}[\mathbb{R}, \mathbb{Z}; +, <]$ -definable
- whether the recognized set is a subsemigroup of $(\mathbb{R}^d, +)$.

Finally, the method given in Section 3.3 is the more efficient method of this section and is the method used in the remaining of this paper. Note that this method leads to proofs which are more complicated than the one needed to apply the two preceding methods.

3.1 Honkala and brute-force algorithm

The method given in this section is based on [9, Theorem 10]. Let \mathcal{L} be a family of regular languages. Conditions about \mathcal{L} are now given. If a class of language satisfies those condition, it is shown that it is decidable whether an automaton recognizes a language belonging to \mathcal{L} .

Let us assume that there exists a size function $s : \mathcal{L} \rightarrow \mathbb{N}$, such that:

1. the number of states of the minimal automaton recognizing a language $L \in \mathcal{L}$ is at least $s(L)$.
2. a finite superset of $s^{-1}(i)$ is computable for all $i \in \mathbb{N}$.

In order to decide whether a minimal automaton \mathcal{A} with n states recognizes a language of \mathcal{L} , it suffices to consider the following algorithm:

- The algorithm runs on each integer $i \in [n]$,
- for each i , the algorithm runs on each language $L \in \mathcal{L}$ such that $s(L) = i$, by Hypothesis (2) it can be done,

- for each L , the algorithm constructs the minimal automaton \mathcal{A}_L which recognizes L ,
- if the minimal automata of \mathcal{A} and of \mathcal{A}_L are equal, the algorithm accepts.

If the algorithm has not accepted, it rejects.

The following proposition shows that this method can be applied to the problem considered in this paper. Note that the algorithm given in the following proof is inefficient.

Proposition 3.1. *It is decidable whether an automaton recognizes a $\text{FO}[\mathbb{R}, \mathbb{Z}; +, <, 1]$ -definable set of reals.*

The proposition also holds for the more general notion of automata recognizing set of tuples of reals, as defined in [3].

Proof. Let $r \in \mathbb{Q}$ and $u.vw^\omega$ be one of its encoding in base b with $|v|$ and $|w|$ minimal. Let the pre-periodic length of r be $|u| + 1 + |v|$ and let the periodic length of r be $|w|$.

Let \mathcal{L} be the class languages which are encodings of simple sets. Let us first consider Hypothesis 1.

Let R be a $\text{FO}[\mathbb{R}, \mathbb{Z}; +, <, 1]$ -definable of real. By [17, Section 2], this logic admits quantifier elimination. Let $\phi \in \Sigma_0[\mathbb{R}, \mathbb{Z}; +, <, 1]$ be a quantifier-free formula defining R . It can be proven that the number of state of the minimal automaton recognizing a simple set R is, at least, the maximum of the pre-periodic lengths and of the periodic lengths of the constant appearing in ϕ . Let $s(\phi)$ be the greatest pre-periodic or periodic length of the constants of ϕ and let $s(R)$ be the minimal $s(\phi)$ for any $\phi \in \Sigma_0[\mathbb{R}, \mathbb{Z}; +, <, 1]$ which defines R .

Let us now consider hypothesis 2, let us give an algorithm which takes as input an integer $i \in \mathbb{N}$ and generates all sets R such that $s(R) = i$. It suffices to remark that, for each $i \in \mathbb{N}$, there is a finite number of rationals whose pre-periodic length and whose periodic length is less than i , let $S^{\leq i}$ be the set of those rationals. The algorithm runs on each subset U of $S^{\leq i}$. For each U the algorithm generates all $\Sigma_0[\mathbb{R}, \mathbb{Z}; +, <, 1]$ -formulas in conjunctive normal form whose rationals belongs to $S^{\leq i}$. Note that there is only a finite number of such formula up to permutation of the elements of the conjunctions. Those formulas define all of the sets R such that $s(R) = i$. \square

3.2 Muchnik and decidable logic

The algorithm given in this section is based on [14, Theorem 3].

Let \mathfrak{S} be a set of subset of $(\mathbb{R}^{\geq 0})^d$ be a set of d -tuples of non-negative reals. Let R be a d -ary symbol representing a set of d -tuples of reals. Let us assume that there exists a $\text{FO}[\mathbb{R}, \mathbb{Z}; X_b, +, <, 1, R]$ -formula ϕ which characterizes whether a set R belongs to \mathfrak{S} . It is decidable whether a RVA \mathcal{A} is such that $[\mathcal{A}]_b^R \in \mathfrak{S}$. Indeed, the equivalence between $\text{FO}[\mathbb{R}, \mathbb{Z}; X_b, +, <, 1]$ and the

Büchi automata is effective, hence it suffices to translate the formula ϕ into a Büchi automaton, according to the algorithm of [4], where R is encoded by the automaton \mathcal{A} .

Let us give an example of application of this method.

Proposition 3.2. *It is decidable whether a Büchi automaton recognizes a set which is a subsemigroup of $(\mathbb{R}^d, +)$.*

Proof. Using the argument given above, it suffices to use the formula:

$$\forall x_0, \dots, x_{d-1}, y_0, \dots, y_{d-1}. [(x_0, \dots, x_{d-1}) \in R \wedge (y_0, \dots, y_{d-1}) \in R] \implies (x_0 + y_0, \dots, x_{d-1} + y_{d-1}) \in R.$$

□

Note that the method introduced in this section leads to inefficient algorithm. The computation time of those algorithms are, at least, a tower of exponential, whose height is equal to the number of quantifier alternation.

3.3 The efficient method

The method introduced in this section leads to proofs which are more difficult than the method given in the two preceding sections. However, this method also leads to more efficient algorithms. The method introduced in this section is the one used in Theorems 4.4 and 5.5. This method is similar to the proofs used in [11] and in [13].

Proposition 3.3. *Let \mathbb{L}' be a class of language and \mathbb{A}' be a class of weak Büchi automata such that $L \in \mathbb{L}'$ if and only if it is recognized by a Büchi automaton of \mathbb{A}' .*

Let \mathbb{L} be a class of languages over an alphabet such that there exists a class \mathbb{A} of weak Büchi automata such that:

1. *it is decidable in time $t(n, b)$ whether a Büchi automaton belongs to \mathbb{A} , for n the number of states and b the number of letters,*
2. *for each $L \in \mathbb{L} \cap \mathbb{L}'$, there exists an automaton $\mathcal{A} \in \mathbb{A}$ which recognizes L ,*
3. *the minimal quotient of any automaton of \mathbb{A} belongs to \mathbb{A} and*
4. *every language recognized by some automaton of \mathbb{A} belongs to \mathbb{L} .*

There exists an algorithm α , which halts in time $O(t(n, b))$, and decides whether a minimal automaton of \mathbb{A}' recognizes a language of \mathbb{L} . Furthermore, the algorithm α applied to an automaton belonging to $\mathbb{A}' \setminus \mathbb{A}$ may return a false negative but it may not return any false positive.

In this paper, \mathbb{A}' is either the set of FNA or of RNA. Considering the class \mathbb{A}' allows to restrict the kind of automata studied. The proposition still hold when "weak Büchi automata" is replaced by "finite automaton". More generally, a

similar proposition can be given as soon as, for each language, there exists a canonical automaton recognizing this language. This requirement is the reason for which this proposition does not hold for non-weak Büchi automata or for non-deterministic automata.

Proof. By Property (1) there exists an algorithm α which accepts in time $t(n, l)$ the automata of \mathbb{A} . Let \mathcal{A} be a weak Büchi automaton and let $L = L_\omega(\mathcal{A})$. Let us prove that if α accepts \mathcal{A} then $L \in \mathbb{L}$ and if $L \in \mathbb{L}$ and $\mathcal{A} \in \mathbb{A}'$ then α accepts \mathcal{A} .

Let us first prove that if α accepts \mathcal{A} , then $L \in \mathbb{L}$. If \mathcal{A} is accepted by α , by Property (1), $\mathcal{A} \in \mathbb{A}$. By Property (4), since $\mathcal{A} \in \mathbb{A}$, $L \in \mathbb{L}$.

Let us now prove that if $L \in \mathbb{L}$ and $\mathcal{A} \in \mathbb{A}'$ then α accepts \mathcal{A} . Since $\mathcal{A} \in \mathbb{A}'$, by definition of \mathbb{L}' , $L \in \mathbb{L}'$. Since $L \in \mathbb{L}$ and $L \in \mathbb{L}'$, by Property (2), there exists an automaton $\mathcal{A}' \in \mathbb{A}$ which recognizes L . Since \mathcal{A}' and \mathcal{A} recognize the same languages and \mathcal{A} is minimal, then \mathcal{A} is the minimal quotient of \mathcal{A}' . Since \mathcal{A}' is the minimal quotient of an automaton belonging to \mathbb{A} , by Property (3) $\mathcal{A} \in \mathbb{A}$. Since $\mathcal{A} \in \mathbb{A}$, by Property (1), the algorithm α accepts the automaton \mathcal{A} . \square

4 Automata accepting simple subsets of $[0, 1]$

This section deals with automata recognizing simple subsets of $[0, 1]$. For each simple set R , a weak Büchi automaton which recognizes $L_\omega(R)$ is defined in Section 4.1. In Section 4.2, an algorithm is given, which accepts the weak FNAs which recognize simple sets.

4.1 From sets to automata

Let $R \subseteq [0, 1]$ be a simple set, in this section, it is explained how to compute an automaton \mathcal{A}_R which recognizes $L_\omega(R)$. This automaton has the form described in [?, Chapter 5]. As seen in Section 1.4.1, R can be expressed as:

$$\bigcup_{i=0}^{I-1} (\rho_{i,\mathfrak{L}}, \rho_{i,\mathfrak{R}}) \cup \bigcup_{i=0}^{J-1} \{\rho_{i,\mathfrak{S}}\}$$

with $\rho_{i,j} \in \mathbb{Q}$. In this section, it can furthermore be assumed that the $\rho_{i,j}$'s belong to $[0, 1]$.

Let $w_{i,j,k}$, with k taking value in $\{0, (b-1)\}$, be the one or two encodings in base b of $\rho_{i,j}$. Let l be an integer such that, for all i, j, k, i', j', k' , either $w_{i,j,k} = w_{i',j',k'}$ or $w_{i,j,k} [< l] \neq w_{i',j',k'} [< l]$. By an easy induction on the number of words $w_{i,j,k}$, such an integer l exists.

Since the $\rho_{i,j}$ are rationals, their encodings in base b are of the form $u_{i,j,k} v_{i,j,k}^\omega$ with $u_{i,j,k} \in \Sigma_b^*$, $v_{i,j,k} \in \Sigma_b^+$ and $k \in \{0, (b-1)\}$. Without loss of generality, let us assume that all $u_{i,j,k} > l$ and that $|v_{i,j,k}|$ is minimal. Those two assumptions imply that no $u_{i,j,k}$ is the prefix of a $u_{i',j',k'}$ and that if $u_{i,j,k} = u_{i',j',k'}$ then $v_{i,j,k} = v_{i',j',k'}$.

A weak FNA \mathcal{A}_R which recognizes $L_\omega(R)$ is now defined.

Definition 4.1 (\mathcal{A}_R). For R a simple set, let $\mathcal{A}_R = (Q_R, \Sigma_b, \delta_R, q_\epsilon, F_R)$ where:

- The set of states Q_R is defined as the set of states q_w for w a strict prefixes of $u_{i,j,k}v_{i,j,k}$, plus two states $q_{\emptyset,R}$ and $q_{[0,1],R}$.
- The transition function is

$$\delta_R(q, a) = \begin{cases} q & \text{if } q \in \{q_{\emptyset,R}, q_{[0,1],R}\}, \\ wa & \text{if } q_w = q \text{ and } wa \text{ is a strict prefix of some } u_{i,j,k}v_{i,j,k}, \\ u_{i,j,k} & \text{if } q = q_w \text{ and } wa = u_{i,j,k}v_{i,j,k}, \\ q_{[0,1],R} & \text{if } q = q_w \text{ and } [wa]_b^R \text{ belongs to an interval } (\rho_{i,\mathfrak{L}}, \rho_{i,\mathfrak{R}}) \\ q_{\emptyset,R} & \text{otherwise.} \end{cases}$$

- The set F_R of accepting states contains $q_{[0,1],R}$ and the states q_w for w some strict prefix of $u_{i,\mathfrak{S},k}v_{i,\mathfrak{S},k}$ of length at least $|u_{i,\mathfrak{S},k}|$, for some i and k .

The last definition is used to characterize the automaton, not to compute it, hence there is no need to study how to decide whether $[wa]_b^R$ belongs to an interval $(\rho_{i,\mathfrak{L}}, \rho_{i,\mathfrak{R}})$, nor to compute the words $u_{i,j,k}$. Note that the name of the states $q_{\emptyset,R}$ and $q_{[0,1],R}$ are consistent with Section 2.

An automaton \mathcal{A}_R and its minimal quotient are now given. The set R is not the same as the one of Example 1.17 because a subset of $[0, 1]$ is now required.

Example 4.2. Let $R = [\frac{1}{4}, \frac{1}{3}) \cup \{\frac{11}{24}, \frac{2}{3}\}$ as in Example 1.17. The following table gives the values associated to those indexes.

$\rho_{i,j}$	its encodings in base 2	$u_{i,j,k}$	$v_{i,j,k}$	$\rho_{i,j}$	its encodings in base 2	$u_{i,j,k}$	$v_{i,j,k}$
$\frac{1}{4}$	001^ω	001	1	$\frac{1}{3}$	$(01)^\omega$	0101	01
$\frac{1}{4}$	010^ω	0100	0	$\frac{2}{3}$	$(10)^\omega$	101	01
$\frac{5}{12}$	$01(10)^\omega$	011	01				

The automaton \mathcal{A}_R is pictured in Figure 4, without the state $q_{\emptyset,R}$, and its minimal quotient is pictured in Figure 4.

Intuitively, the states q_1 and q_{10} are used to read the binary encoding of $\frac{2}{3}$. The states q_{00} and q_{001} are used to read one of the binary encoding of $\frac{1}{4}$ and q_{0100} is used to read its other encoding. The states q_{011} and q_{0111} are used to read $\frac{11}{24}$. Finally, the states q_{0101} and q_{01010} are used to read $\frac{1}{3}$.

Note that minimization sends the states q_{011} and q_{10} to the same state, named q_{10} . Intuitively, it is because when an automaton reads a rational r belonging to the boundary of R , the only important information is:

- the periodic part of the encoding of r ,
- whether r belongs to R and

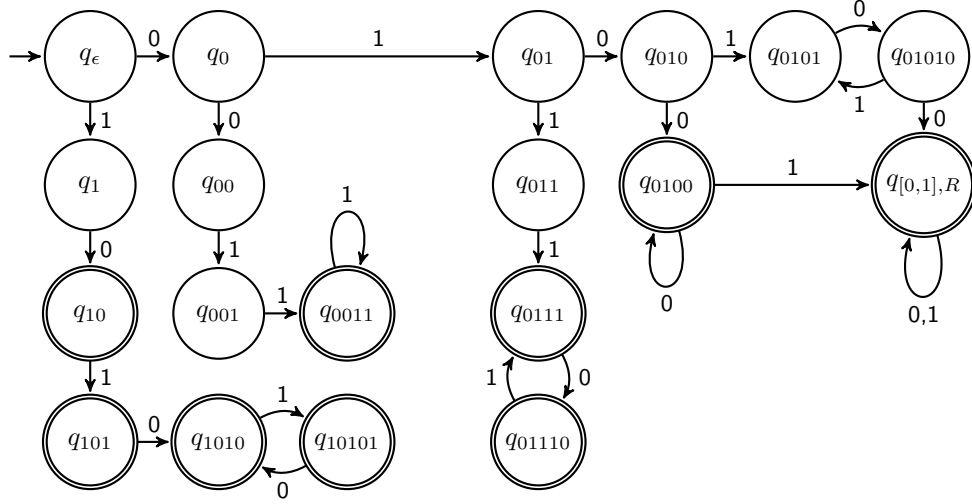


Figure 4: The automaton \mathcal{A}_R for $R = [\frac{1}{4}, \frac{1}{3}) \cup \{\frac{11}{24}, \frac{2}{3}\}$

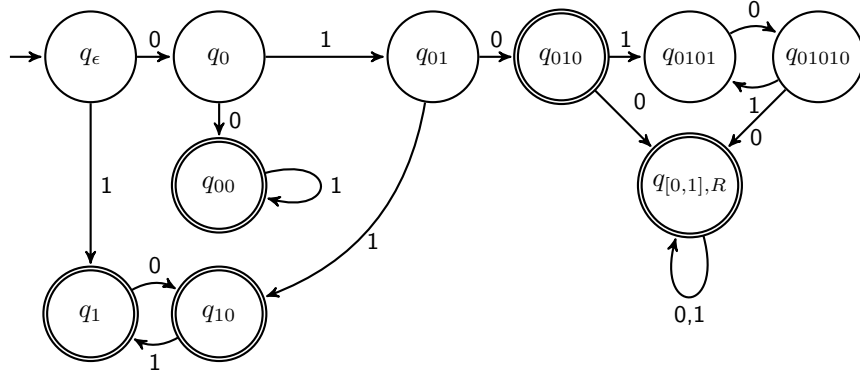


Figure 5: The minimal automaton recognizing $R = [\frac{1}{4}, \frac{1}{3}) \cup \{\frac{11}{24}, \frac{2}{3}\}$

- whether, $(r - \epsilon, r)$ (respectively, $(r, r + \epsilon)$) is included in R or is disjoint from R , for ϵ small enough.

Let us prove that \mathcal{A}_R is as expected.

Lemma 4.3. *Let $R \subseteq [0, 1]$ be a simple set. The automaton \mathcal{A}_R recognizes $L_\omega(R)$.*

Proof. Let $w \in \Sigma_b^\omega$, and let $r = [w]_b^R$. Let us prove that $[w]_b^R \in R$ if and only if w is accepted by \mathcal{A}_R . Two cases must be considered, depending on whether r is equal to some $\rho_{i,\mathfrak{S}}$ for some i , or not. In the second cases, four more cases must be considered, depending on whether r is equal to some $\rho_{i,\mathfrak{L}}$, whether r is equal

to some $\rho_{i,\mathfrak{A}}$, whether r belongs to some $(\rho_{i,\mathfrak{L}}, \rho_{i,\mathfrak{A}})$ for some i , or whether r does not satisfy any of those properties.

- Let us first assume that $r = \rho_{i,\mathfrak{S}}$ for some i . Then $r \in R$. Let us show that w is accepted by \mathcal{A}_R . Since $r = \rho_{i,\mathfrak{S}}$, then w is of the form $u_{i,\mathfrak{S},k} v_{i,\mathfrak{S},k}^\omega$, hence, by induction on the prefixes of w , all visited states are of the form q_w for w some prefix of $u_{i,\mathfrak{S},k} v_{i,\mathfrak{S},k}$. It implies that each state of the run of \mathcal{A}_R on w – apart from the $|u_{i,\mathfrak{S},k}|$ first states – are accepting. Hence \mathcal{A}_R accepts w .
- From now on, let us assume that, for all i , $r \neq \rho_{i,\mathfrak{S}}$. Let us suppose that $r = \rho_{i,\mathfrak{L}}$ for some i . The case $r = \rho_{i,\mathfrak{R}}$ is similar. In this case, $r \notin R$. Let us prove that w is not accepted by \mathcal{A}_R . The word w is of the form $u_{i,j,k} v_{i,j,k}^\omega$, hence by induction on the prefix's of w , each visited state is q_w for w a prefix of $u_{i,j,k} v_{i,j,k}$, and since the $u_{i,j,k}$ are not prefixes of $u_{i',j',k'} v_{i',j',k'}^\omega$, then no accepting state is visited. Hence \mathcal{A}_R does not accepts w .
- Let us assume that $r \in (\rho_{i,\mathfrak{L}}, \rho_{i,\mathfrak{A}})$. Let us show that \mathcal{A}_R accepts w . Three cases must be considered, depending on whether $u_{i,\mathfrak{L},k}$ is a prefix of w for some $k \in \{0, (b-1)\}$, whether $u_{i,\mathfrak{A},k}$ is a prefix of w for some $k \in \{0, (b-1)\}$, or whether none of those words are prefix of w .

- Let us suppose that $u_{i,\mathfrak{L},k}$ is a prefix of w , then there exists a unique 4-tuple $n \in \mathbb{N}$, $h < |v_{i,\mathfrak{L},k}|$, $a < v_{i,\mathfrak{L},k}[h]$ and $w' \in \Sigma_b^\omega$ such that

$$w = u_{i,\mathfrak{L},k} v_{i,\mathfrak{L},k}^n (v_{i,\mathfrak{L},k} [< h]) a w'.$$

Since $a < v_{i,\mathfrak{L},k}[h]$, $[u_{i,\mathfrak{L},k} v_{i,\mathfrak{L},k} [< h] a w']_b^R > \rho_{i,\mathfrak{L}}$. Since the $u_{i,j,k}$'s are not prefixes of the $u_{i',j',k'}$'s, $[u_{i,\mathfrak{L},k} v_{i,\mathfrak{L},k} [< h] a w']_b^R < \rho_{i,\mathfrak{L}}$. Then, for $i > \left| u_{i,\mathfrak{L},k} v_{i,\mathfrak{L},k}^n v_{i,\mathfrak{L},k} [< h] \right|$, the i th state is $q_{[0,1],R}$, hence w is accepted by R .

- If $u_{i,\mathfrak{A},k}$ is a prefix of w , the proof is similar by symmetry.
- Let us now assume that there are no j and k such that $u_{i,j,k}$ is a prefix of w . Let us consider the longest prefix p of w and $u_{i,j,k}$. The $|p|$ first steps of the run of \mathcal{A}_R on w are the states q_v for v some prefix of p . The following steps are $q_{[0,1],R}$, hence \mathcal{A}_R accepts w .
- In all other cases, the proof is similar to the preceding case, replacing $q_{[0,1],R}$ by $q_{\emptyset,R}$.

□

4.2 Characterization of automata recognizing simple sets

The main theorem of this paper concerning subsets of $[0, 1]$ is now stated.

Theorem 4.4. *It is decidable in time $O(nb)$ whether a minimal FNA over the alphabet Σ_b with n states recognizes a simple set.*

The proof of this theorem uses Proposition 3.3. In order to use this proposition, a set $\mathbb{A}_{F,S}$ of automata is now introduced. Four lemmas are then proved, which corresponds to the four properties of Proposition 3.3.

Definition 4.5 ($\mathbb{A}_{F,S}$). Let $\mathbb{A}_{F,S}$ be the set of weak Büchi automata $\mathcal{A} = (Q, \Sigma_b, \delta, q_0, F)$ such that, for each strongly connected component $C \subseteq Q_{F,\mathcal{A}} \setminus (Q_{[0,1],\mathcal{A}} \cup Q_{\emptyset,\mathcal{A}})$, there exists $\beta_{<,C}$ and $\beta_{>,C}$, two states of $Q_{[0,1],\mathcal{A}} \cup Q_{\emptyset,\mathcal{A}}$, such that, for all $q \in C$:

1. C is a cycle,
2. for all $a > a_q$, $\delta(q, a)$ is $\beta_{>,C}$ and
3. for all $a < a_q$, $\delta(q, a)$ is $\beta_{<,C}$.

Property (2) implies that $Q_{[0,1],\mathcal{A}} \cup Q_{\emptyset,\mathcal{A}}$ is not empty.

Example 4.2 is resumed in order to show the construction of the preceding lemma.

Example 4.6. Let $R = (\frac{1}{3}, 2] \cup (\frac{8}{3}, 3] \cup (\frac{11}{3}, \infty]$ and \mathcal{A} be the minimal automaton of Figure 5. Let us first consider the recurrent states q equal to 1, 00, 10, 0101 or to 01010. The integer n , the sequence of letters a_0, \dots, a_{n-1} , the states $\beta_{<,C}$ and $\beta_{>,C}$ associated to each of those states q are given in the following table and the Boolean $\beta_{=,C}$ which is true if and only if C is composed of accepting states. The two last columns of the table show the language of infinite words recognized by the automaton \mathcal{A}_q , and the set of reals recognized by this state.

q	n	a_0	a_1	$\beta_{<,C}$	$\beta_{=,C}$	$\beta_{>,C}$	$L_\omega(\mathcal{A}_q)$	$[\mathcal{A}_q]_b^R$
1	2	0	1	$q\emptyset, \mathcal{A}$	True	$q\emptyset, \mathcal{A}$	$(01)^\omega$	$\{\frac{1}{3}\}$
10	2	1	0	$q\emptyset, \mathcal{A}$	True	$q\emptyset, \mathcal{A}$	$(10)^\omega$	$\{\frac{2}{3}\}$
00	1	1		$q\emptyset, \mathcal{A}$	True	$q\emptyset, \mathcal{A}$	$(1)^\omega$	$\{1\}$
0101	2	0	1	$q_{[0,1],\mathcal{A}}$	False	$q\emptyset, \mathcal{A}$	$0(10)^\omega + (10)^*0(0+1)^\omega$	$[0, \frac{1}{3})$
01010	2	1	0	$q_{[0,1],\mathcal{A}}$	False	$q\emptyset, \mathcal{A}$	$(10)^\omega + (10)^*0(0+1)^\omega$	$[0, \frac{2}{3})$
$q_{[0,1],R}$							$(0+1)^\omega$	$[0, 1]$
$q_{\emptyset,R}$							\emptyset	\emptyset

Transient states 010, 01, 0 and ϵ are now considered. One has:

$$\begin{aligned}
[\mathcal{A}_{010}]_b^R &= \frac{0+[\mathcal{A}_{q_{[0,1],R}}]_b^R}{2} \cup \frac{1+[\mathcal{A}_{0101}]_b^R}{2} = \frac{0+[0,1]}{2} \cup \frac{1+[0,\frac{1}{3}]}{2} = [0, \frac{2}{3}), \\
[\mathcal{A}_{01}]_b^R &= \frac{0+[\mathcal{A}_{010}]_b^R}{2} \cup \frac{1+[\mathcal{A}_{10}]_b^R}{2} = \frac{0+[0,\frac{2}{3})}{2} \cup \frac{1+\{\frac{2}{3}\}}{2} = [0, \frac{1}{3}) \cup \{\frac{5}{6}\}, \\
[\mathcal{A}_0]_b^R &= \frac{0+[\mathcal{A}_{00}]_b^R}{2} \cup \frac{1+[\mathcal{A}_{01}]_b^R}{2} = \frac{0+\{1\}}{2} \cup \frac{1+\{[0,\frac{1}{3}) \cup \{\frac{5}{6}\}\}}{2} = [\frac{1}{2}, \frac{2}{3}) \cup \{\frac{11}{12}\}, \\
[\mathcal{A}_\epsilon]_b^R &= \frac{0+[\mathcal{A}_0]_b^R}{2} \cup \frac{1+[\mathcal{A}_1]_b^R}{2} = \frac{0+[\frac{1}{2}, \frac{2}{3}) \cup \{\frac{11}{12}\}]}{2} \cup \frac{1+\{\frac{1}{3}\}}{2} = (\frac{1}{3}, 2] \cup (\frac{8}{3}, 3] \cup (\frac{11}{3}, \infty].
\end{aligned}$$

It is now proven that the set $\mathbb{A}_{F,S}$ satisfies Property (1) of Proposition 3.3.

Lemma 4.7. *It is decidable in time $O(nb)$ whether a weak Büchi automaton with n states belongs to $\mathbb{A}_{F,S}$.*

Proof. Tarjan's algorithm [16] can be used to compute the set of strongly connected component in time $O(nb)$ and thus the set of recurrent state. By Lemma 2.5, the sets $Q_{\emptyset, \mathcal{A}}$ and $Q_{[0,1], \mathcal{A}}$ are computable in linear time. The algorithm runs on each strongly connected component distinct from $Q_{[0,1], \mathcal{A}}$ and from $Q_{\emptyset, \mathcal{A}}$.

It is now explained how the algorithm checks whether Property (1) is satisfied by the automaton. The algorithm runs on each $q \in Q$. The algorithm keeps a counter c_q , initialized to 0, of the number of letters $a \in \Sigma_b$ such that $\delta(q, a) \in C$. For each q , the algorithm runs on each letter $a \in \Sigma_b$. For each a , the algorithm tests whether $\delta(q, a) \in C$, and if it is the cases, c_q is incremented. If $c_q \neq 1$ the algorithm rejects.

It is now explained how the algorithm checks whether Property (2) is satisfied by the automaton. Checking Property (3) is done similarly. The algorithm runs on each $q \in C$. If $a_q > 0$ then:

- if $\beta_{<,C}$ is not set, then $\beta_{<,C}$ is set to $\delta(q, 0)$.
- otherwise, let us assume that $\beta_{<,C}$ is set. The algorithm runs on each $0 \leq a < a_q$. For each a , if $\delta(q, a)$ is different from $\beta_{<,C}$, then the algorithm rejects.

If the algorithm has not rejected, it accepts. \square

It is now proven that the set $\mathbb{A}_{F,S}$ satisfies Property (2) of Proposition 3.3.

Lemma 4.8. *Let $R \subseteq [0, 1]$ be a simple set. The automaton \mathcal{A}_R belongs to $\mathbb{A}_{F,S}$.*

Proof. It suffices to see that the recurrent states $q \notin Q_{\emptyset, \mathcal{A}} \cup Q_{[0,1], \mathcal{A}}$ of \mathcal{A}_R are of the form $u_{i,j,k}w$ with w a prefix of $v_{i,j,k}$. For each prefix w of length l of $v_{i,j,k}$, the digit $s_{q_{u_{i,j,k}w}}$ is $v_{i,j,k}[l]$, thus Property (1) holds. The state $\beta_{<,C}$ (respectively, $\beta_{>,C}$) is $q_{\emptyset, \mathcal{A}}$ if $u_{i,j,k}0^\omega \notin R$ and $q_{[0,1], \mathcal{A}}$ otherwise, thus Property (2) (respectively 3) holds. Then the conditions of Definition 4.5 are satisfied. \square

It is now proven that the set $\mathbb{A}_{F,S}$ satisfies Property (3) of Proposition 3.3.

Lemma 4.9. *The minimal quotients of automata of $\mathbb{A}_{F,S}$ belong to $\mathbb{A}_{F,S}$.*

Proof. Let $\mathcal{A} = (Q, \Sigma_b, \delta, q_0, F)$ be an automaton belonging to $\mathbb{A}_{F,S}$ and let its minimal quotient be $\mathcal{A}' = (Q', \Sigma_b, \delta', q'_0, F')$. Let μ be the morphism from \mathcal{A} to \mathcal{A}' . Let us show that \mathcal{A}' belongs to $\mathbb{A}_{F,S}$. Let C' be a component included in \mathcal{A}' , distinct from $Q_{\emptyset, \mathcal{A}'}$ and from $Q_{[0,1], \mathcal{A}'}$. By Lemma 1.14, there exists a strongly connected component C of \mathcal{A} such that $\mu(C) = C'$ and such that, for all $q \in Q \setminus C$ accessible from C , $\mu(q) \notin C'$.

Let us first show that Property (1) is satisfied by \mathcal{A}' . Let $q' \in C'$ and let us prove that there exists exactly one letter $s_{q'}$ such that $\delta'(q', s_{q'}) \in C'$. Since q'

is recurrent, at least one such letter exists. It remains to prove the unicity. Let us assume that there exists two letters a_0 and a_1 such that $\delta'(q', a_0) \in C'$ and $\delta'(q', a_1) \in C'$ and let us prove that $a_0 = a_1$. Since $\mu(C) = C'$, there exists $q \in C$ such that $\mu(q) = q'$. Since $\mathcal{A} \in \mathbb{A}_{F,S}$, there exists exactly one letter s_q such that $\delta(q, s_q) \in C$. It suffices to prove that $a_0 = s_q = a_1$. Let us prove that $a_0 = s_q$, the other case is similar. Since $\mu(\delta(q, a_0)) = \delta'(\mu(q), a_0) = \delta'(q', a_0) \in C'$ and since $\delta(q, a_0)$ is accessible from C , by hypothesis about C , $\delta(q, a_0) \in C$. Since $\delta(q, a_0) \in C$, by Property (1) applied to \mathcal{A} , by definition of s_q , $a_0 = s_q$.

It is now proven that \mathcal{A}' satisfies Property (2). The case of Property (3) is similar. Let q'_1 and q'_2 be two states of C' , and let $a_1 > s_{q'_1}$ and $a_2 > s_{q'_2}$. Since $\mu(C) = C'$, there exists $q_1, q_2 \in C$ such that $\mu(q_1) = q'_1$ and $\mu(q_2) = q'_2$. Let q_1 and q_2 be those two antecedents, note that $s_{q_1} = s_{q'_1}$ and $s_{q_2} = s_{q'_2}$. Since $\mathcal{A} \in \mathbb{A}_{F,S}$, by Property applied to \mathcal{A} 2 $\delta(q_1, a_1) = \delta(q_2, a_2)$, hence $\delta'(q'_1, a_1) = \delta'(q'_2, a_2)$. \square

It is now proven that the set $\mathbb{A}_{F,S}$ satisfies Property (4) of Proposition 3.3.

Lemma 4.10. *The automata of $\mathbb{A}_{F,S}$ recognize simple subsets of $[0, 1]$.*

Note that it is not required in this lemma that the Büchi automaton is a FNA. In order to prove this lemma, another lemma is now required. Its proof is straightforward from Definition 4.5.

Lemma 4.11. *Let \mathcal{A} be a Büchi automaton of $\mathbb{A}_{F,S}$ and let q be a state of \mathcal{A} . The automaton \mathcal{A}_q belongs to $\mathbb{A}_{F,S}$.*

Proof of Lemma 4.10. The proof is by induction on the number n of states of \mathcal{A} . Two cases must be considered, depending on whether the initial state of the automaton is recurrent or not.

Let us first suppose that the initial state is transient. For each $a \in \Sigma_b$, by Lemma 4.11, the automaton $\mathcal{A}_{\delta(q_0, a)}$ belongs to $\mathbb{A}_{F,S}$. The automaton $\mathcal{A}_{\delta(q_0, a)}$ has less than n states and belongs to $\mathbb{A}_{F,S}$, hence, by induction hypothesis, it recognizes a set R_a of the form $\bigcup_{i=0}^{I_a-1} (\rho_{i, \mathfrak{L}, a}, \rho_{i, \mathfrak{R}, a}) \cup \bigcup_{i=0}^{J_a-1} \{\rho_{i, \mathfrak{S}, a}\}$. The set $[\mathcal{A}]_b^R$ is then equal to:

$$\bigcup_{a=0}^{b-1} \frac{a + R_a}{b} = \bigcup_{a=0}^{b-1} \bigcup_{i=0}^{I_a-1} \left(\frac{a + \rho_{i, \mathfrak{L}, a}}{b}, \frac{a + \rho_{i, \mathfrak{R}, a}}{b} \right) \cup \bigcup_{i=0}^{J_a-1} \left\{ \frac{a + \rho_{i, \mathfrak{S}, a}}{b} \right\},$$

where the $\rho_{i,j}$ belongs to $\mathbb{Q} \cap [0, 1]$. Hence $[\mathcal{A}]_b^R$ is a simple set.

Let us now assume that the initial state is recurrent. Three cases must be considered depending on whether q_0 belongs to $Q_{\emptyset, \mathcal{A}}$, to $Q_{[0,1], \mathcal{A}}$ or to neither of those sets. Let us first assume that $q_0 \in Q_{\emptyset, \mathcal{A}}$, the case $q_0 \in Q_{\infty, \mathcal{A}}$ is similar. The automaton recognizes \emptyset by definition of $Q_{\emptyset, \mathcal{A}}$, thus is simple.

Let us suppose that $q_0 \notin Q_{\emptyset, \mathcal{A}} \cup Q_{[0,1], \mathcal{A}}$ and let C be the strongly connected component of q_0 . By Property (1) of Definition 4.5, there exists a sequence a_0, \dots, a_{n-1} such that $\delta(q_0, a_0 \dots a_{n-1}) = q_0$. Let c_{q_0} be the real represented

by $(a_0 \dots a_{n'-1})^\omega$ that is $\frac{\sum_{i=0}^{n'-1} a_i b^{n'-i-1}}{b^{n'} - 1}$. The number c_{q_0} is rational, and its length is $O(\log(b)n)$. Let $w \in \Sigma_b^\omega$ and $x = [w]_b^R$. It is then clear from Definition 4.5, that \mathcal{A} accepts w if and only if, either $x = c_{q_0}$ and $C \subseteq F$, either $x < c_{q_0}$ and $\beta_{<,C} \in Q_{F,\mathcal{A}}$, or similarly $x > c_{q_0}$ and $\beta_{>,C} \in Q_{F,\mathcal{A}}$. \square

Theorem 4.4 can now be proven.

Proof. It suffices to use Proposition 3.3 with \mathbb{A} being the set of automata $\mathbb{A}_{F,S}$, \mathbb{A}' being the set of FNA, and Lemmas 4.7, 4.8, 4.9 and 4.10. \square

Note that the algorithm given in the proof of Theorem 4.4 returns no false positive even when it is applied to a Büchi Automaton which is not a FNA. The author conjecture that there exists no false negative.

5 Simple subsets of $\mathbb{R}^{\geq 0}$

In the preceding section, the problem studied in this paper was solved on $[0, 1]$. The general problem is solved in this section using the notations and lemmas of Section 4.

In Section 5.1, given a simple set R , a weak automaton \mathcal{A}_R is constructed, which recognizes R . In Section 5.2, an algorithm is given, which takes as input a weak automaton of alphabet Σ_b^* and decides whether it recognizes a simple set.

5.1 From sets to automata

Let us fix in this section a simple set $R \subseteq \mathbb{R}^{\geq 0}$. Since R is a simple set, there exists a least integer $t_R \in \mathbb{N}^{\geq 0}$ such that $[t_R, \infty)$ is either a subset of R or is disjoint from R . For all $i \in \mathbb{N}$, let R_i denote $\{x \in [0, 1] \mid x + i \in R\}$, and let $\mathcal{A}_i = (Q_i, \Sigma_b, \delta_i, q_{0,i}, F_i)$ be the minimal automaton accepting $[R_i]_b^R$. By Example 1.4 is now resumed.

Example 5.1. Let $R = (\frac{1}{3}, 2] \cup (\frac{8}{3}, 3] \cup (\frac{11}{3}, \infty]$ as in Example 1.4. Then t_R is 4, $R_0 = (\frac{1}{3}, 1]$, $R_1 = [0, 1]$, and $R_2 = R_3 = \{0\} \cup [\frac{2}{3}, 1]$.

In this section, a weak RNA \mathcal{A}_R which recognizes $L_\omega(R)$ is constructed. The part of \mathcal{A}_R which reads the fractional parts of the reals is based on the construction of Section 4.1. The formal definition of \mathcal{A}_R is now given.

Definition 5.2 (\mathcal{A}_R). Let $R \subsetneq [0, \infty)$ be a simple non-empty set. Note that $t_R > 0$. Let \mathcal{A}_R be the automaton $(Q, \Sigma_b^*, \delta, q_0, F)$ where:

- the set Q of state contains:
 - a state q_i for all $i \in [t_R - 1]$,
 - a state $q_{0,R}$,
 - a state $q_{\infty,R}$ if $[T_R, \infty) \subseteq R$,

- a state (i, q) for all $i \in [t_R - 1]$, and for each state $q \in Q_i \setminus q_{\emptyset, \mathcal{A}_i}$.
- The initial state is q_0 .
- The accepting states are $q_{[0,1], \mathcal{A}}$ and (i, q) for all accepting states q of \mathcal{A}_i .
- The transition function is such that, for all $i \in [t_R - 1]$, $a \in \Sigma_b$:
 - $\delta(q, a) = q$ for q being $q_{[0,1], \mathcal{A}}$, $q_{\infty, R}$ or $q_{\emptyset, R}$,
 - $\delta(q, \star) = q_{\emptyset, R}$ for q being $q_{[0,1], \mathcal{A}}$ or $q_{\emptyset, R}$,
 - $\delta(q_{\infty, R}, \star)$ is $q_{[0,1], R}$,
 - $\delta(q_i, a)$ is $bi + a$ if $bi + a < t_R$,
 - $\delta(q_i, a)$ is $q_{\infty, R}$ if $bi + a \geq t_R$ and if $[t_R, \infty) \subseteq R$,
 - $\delta(q_i, a)$ is $q_{\emptyset, R}$ if $bi + a \geq t_R$ and if $[t_R, \infty) \cap R = \emptyset$,
 - $\delta(q_i, \star) = (i, q_{0,i})$,
 - $\delta((i, q), a)$ is $(i, \delta_i(q, a))$ for $q \in \mathcal{A}_i$, if $\delta_i(q, a) \notin Q_{\emptyset, \mathcal{A}_i}$
 - $\delta((i, q), a)$ is $q_{\emptyset, \mathcal{A}}$ for $q \in \mathcal{A}_i$, if $\delta_i(q, a) \in Q_{\emptyset, \mathcal{A}_i}$ and
 - $\delta((i, q), \star)$ is $q_{\emptyset, R}$.

An example of automaton \mathcal{A}_R is now given, resuming Example 5.1.

Example 5.3. Let $R = (\frac{1}{3}, 2] \cup (\frac{8}{3}, 3] \cup (\frac{11}{3}, \infty]$, as in Example 5.1. The automaton \mathcal{A}_R is pictured in Figure 1, without the non accepting state $q_{\emptyset, \mathcal{A}}$. Its minimal quotient is pictured in Figure 2.

Let us now show that \mathcal{A}_R is as expected.

Proposition 5.4. *Let $R \subsetneq \mathbb{R}^{\geq 0}$ be a simple non-empty set. The automaton \mathcal{A}_R recognizes $L_\omega(R)$.*

Proof. Let $w_I \in \Sigma_b^*$, $w_F \in \Sigma_b^\omega$ and $w = w_I \star w_F$. Let $r_I = [w_I]_b^I$, $r_F = [w_F]_b^R$ and $r = [w]_b^R = r_I + r_F$. Let us prove that $[w]_b^R \in R$ if and only if w is accepted by \mathcal{A}_R .

By an easy induction on the length of w_I , $\delta(q_0, w_I)$ is q_{w_I} if $r_I < t_R$, otherwise it is $q_{\infty, R}$ if $[t_R, \infty) \subseteq R$ and it is $q_{\emptyset, \mathcal{A}}$ otherwise. Two cases are considered, depending on whether $r_I < t_R$ or whether $r_I \geq t_R$.

- Let us first assume that $r_I < t_R$. Then $\delta(0, w_I) = q_{w_I}$ and thus $\delta(0, w_I \star) = (r_I, q_{0, r_I})$. By Lemma 4.3, \mathcal{A}_{r_I} recognizes R_{r_I} , hence w_F is accepted by $(\mathcal{A}_R)_{(r_I, q_{0, r_I})}$ if and only if $r_F \in \mathcal{A}_{r_I}$. Hence w is accepted by \mathcal{A}_R if and only if $r \in R$.
- Let us now assume that $r_I \geq t_R$. Let us assume that $[t_R, \infty) \subseteq R$, the case $[t_R, \infty) \cap R = \emptyset$ is similar. Since $r_i \geq t_R$, then $r \in R$. By definition of \mathcal{A}_R , $\delta(q_0, w_I) = q_{\infty, R}$, therefore $\delta(q_0, w_I \star) = q_{[0,1], \mathcal{A}}$. It follows that each state of the run of \mathcal{A}_R on w is $q_{[0,1], \mathcal{A}}$, apart from the $|w_I \star|$ first ones. Since furthermore $q_{[0,1], \mathcal{A}}$ is an accepting state, \mathcal{A}_R accepts w .

□

5.2 Characterization of automata recognizing simple sets

The first main theorem of this paper is now given.

Theorem 5.5. *It is decidable in time $O(nb)$ whether a minimal weak Büchi RNA with n states recognizes a simple set.*

In order to simplify the proof, this theorem is reduced to a simpler case given in the following proposition.

Proposition 5.6. *It is decidable in time $O(nb)$ whether a minimal automaton with n states recognizes a simple set different from \emptyset and from $[0, \infty)$.*

As in Section 4.2, a set of automata $\mathbb{A}_{R \geq 0, s}$ is now introduced. Four lemmas are then proved, which corresponds to the four properties of Proposition 3.3.

Definition 5.7 ($\mathbb{A}_{R \geq 0, s}$). Let $\mathbb{A}_{R \geq 0, s}$ be the set of weak Büchi automata $\mathcal{A} = (Q, \Sigma_b^*, \delta, q_0, F)$ such that:

1. The automaton \mathcal{A} satisfies the properties of Definition 4.5.
2. There exists an accepting and a rejecting strongly connected component, accessible from the initial state, belonging to $Q_{F, \mathcal{A}}$.
3. $\delta(q_0, 0) = q_0$.
4. The set $Q_{\emptyset, \mathcal{A}}$ contains exactly one recurrent state. Let $q_{\emptyset, \mathcal{A}}$ denotes its only state.
5. The set $Q_{\infty, \mathcal{A}}$ contains at most one recurrent element. If $Q_{\infty, \mathcal{A}}$ contains one recurrent element, let $q_{\infty, \mathcal{A}}$ denote this only element.
6. $\delta(q_0, a) \neq q_0$ for all $0 < a < b$.
7. If $q_{\infty, \mathcal{A}}$ exists, then $\delta(q, a) \neq q_{\emptyset, \mathcal{A}}$ for all $q \in Q_{I, \mathcal{A}} \setminus \{q_{\emptyset, \mathcal{A}}\}$ and $a \in \Sigma_b$.
8. Let q be a natural recurrent state. The state q is either $q_{\emptyset, \mathcal{A}}$, $q_{\infty, \mathcal{A}}$ or q_0 .

Note that the properties of Definition 4.5 only consider the states of $Q_{F, \mathcal{A}}$. Therefore, there is no trouble to state that those properties are satisfied on some subset of an automaton closed under the function δ . Let us show that $\mathbb{A}_{R \geq 0, s}$ admits the properties of Proposition 3.3. It is now proven that the set $\mathbb{A}_{R \geq 0, s}$ satisfies Property (1) of Proposition 3.3.

Lemma 5.8. *It is decidable in time $O(nb)$ whether a weak Büchi automaton \mathcal{A} with n states belong to $\mathbb{A}_{R \geq 0, s}$.*

Proof. It suffices to check that all properties of Definition 5.7 are testable in time $O(nb)$. To check Property (1), it suffices to use the algorithm of Lemma 4.7.

The set of states accessible from q_0 is easily computed by a fixed-point algorithm in time $O(nb)$. Using Tarjan's algorithm [16], the set of strongly connected component are computable in time $O(nb)$. By Lemma 2.5, the sets $Q_{\emptyset, \mathcal{A}}$

and $Q_{\infty, \mathcal{A}}$ are computable in time $O(nb)$. It easily follows that testing whether $q_{\infty, \mathcal{A}}$ exists is testable in time $O(nb)$. Since the set of recurrent states, the states $q_{\emptyset, \mathcal{A}}$ and $q_{\infty, \mathcal{A}}$, the set of states accessible from q_0 , the sets $Q_{\emptyset, \mathcal{A}}$, $Q_{\infty, \mathcal{A}}$, $Q_{I, \mathcal{A}}$ and $Q_{F, \mathcal{A}}$ are computed, it is trivial to test the seven last properties in time $O(nb)$. \square

It is now proven that the set $\mathbb{A}_{R \geq 0, s}$ satisfies Property (2) of Proposition 3.3.

Lemma 5.9. *Let $R \subsetneq [0, \infty)$ be a simple non-empty set. The automaton \mathcal{A}_R belongs to $\mathbb{A}_{R \geq 0, s}$.*

Proof. By Lemma 4.8 and Lemma 4.9, all automata \mathcal{A}_i belongs to $\mathbb{A}_{F, S}$. Since Property (1) only consider the states of $Q_{F, \mathcal{A}}$, that is the states of the form (i, q) for $q \in Q_i$, then \mathcal{A}_R satisfies Property (1).

Property (2) is now considered. Since R is neither the empty set nor $[0, \infty)$, there exists a word belonging to $\Sigma_b^* \star \Sigma_b^\omega$ which is rejected by \mathcal{A}_R and a word belonging to $\Sigma_b^* \star \Sigma_b^\omega$ which is accepted by \mathcal{A}_R . Therefore \mathcal{A}_R satisfies Property (2).

Property (3) is now considered. Recall that the threshold of R , t_R is positive. Thus $b0 + 0 < t_R$ and by construction of \mathcal{A}_R , $\delta(q_0, 0) = q_{b0+0} = q_0$, therefore \mathcal{A}_R satisfies 3.

The automaton \mathcal{A}_R satisfy Properties 4, 5 and 7 by construction.

Property (6) is now considered. Let $a > 0$, $q \in Q_{I, \mathcal{A}}$ and let us prove that $\delta(q, a) \neq q_0$. Two cases must be considered, depending on whether $t_R \geq a$ or whether $a < t_R$. Let us first assume that $t_R \geq a$. Note that $b0 + a \geq t_R$. By construction of \mathcal{A}_R , $\delta(q_0, a)$ is either $q_{\infty, R}$ or is $q_{\emptyset, R}$, which are distinct from $q_{q_0, \mathcal{A}}$. Finally, let us assume that $a < t_R$. By construction of \mathcal{A}_R $\delta(q_0, a) = q_{b0+a} = q_a \neq q_0$, therefore \mathcal{A}_R satisfy Property (6).

Property (8) is now considered. By construction of \mathcal{A}_R , there are at most three recurrent states in $Q_{I, \mathcal{A}}$: the initial state, the state $q_{\emptyset, R} \in Q_{\emptyset, \mathcal{A}_R}$, and potentially $q_{\infty, R} \in Q_{\infty, \mathcal{A}_R}$. Therefore \mathcal{A}_R satisfy Property (8). \square

It is now proven that the set $\mathbb{A}_{R \geq 0, s}$ satisfies Property (3) of Proposition 3.3.

Lemma 5.10. *The minimal quotient of automata of $\mathbb{A}_{R \geq 0, s}$ belong to $\mathbb{A}_{R \geq 0, s}$.*

Proof. Let $\mathcal{A} = (Q, \Sigma_b^*, \delta, q_0, F)$ belonging to $\mathbb{A}_{R \geq 0, s}$ and let $\mathcal{A}' = (Q', \Sigma_b^*, \delta', q'_0, F')$ be a quotient of \mathcal{A} by a morphism μ . It is now shown that \mathcal{A}' belongs to $\mathbb{A}_{R \geq 0, s}$. Each of the 8 propositions of Definition 5.7 are considered separately.

Property (1) is first considered. By Lemma 4.9, the set of automata satisfying Property (1) is closed under quotient. Since \mathcal{A} satisfies Property (1), and since \mathcal{A}' is a quotient of \mathcal{A} , \mathcal{A}' satisfies Property (1).

Property (2) is now considered. It is now proven that there exists an accepting recurrent state q' accessible from the initial state of \mathcal{A}' . The case of a non-accepting state is similar. By Property (2), there exists an accepting recurrent state q accessible from q_0 , in $Q_{F, \mathcal{A}}$. Since q is accessible from q_0 , there exists a finite word such that $\delta(q_0, w) = q$. Since q is a recurrent, there exists a non-empty word v such that $\delta(q, v) = q$. Since q is recurrent and accepting, it does

not belongs to $Q_{\emptyset, \mathcal{A}}$. Since q is fractional, since \mathcal{A} accepts a subset of $\Sigma_b^* \star \Sigma_b^\omega$, and since $q \notin Q_{\emptyset, \mathcal{A}}$, then $w \in \Sigma_b^* \star \Sigma_b^*$. Note that $\delta'(\mu(q), v) = \mu(\delta(q, v)) = \mu(q)$, therefore $\mu(q)$ is recurrent. By Lemma 2.7, $\mu(q)$ is fractional. Furthermore $\delta'(q'_0, w) = \delta'(\mu(q_0), w) = \mu(\delta(q_0, w)) = \mu(q)$, therefore q is accessible from q'_0 . Since q is an accepting recurrent state of $Q_{F, \mathcal{A}'}$, accessible from q'_0 , \mathcal{A}' satisfies Property (2).

Property (3) is now considered. Since $\delta'(q'_0, 0) = \delta'(\mu(q_0), 0) = \mu(\delta(q_0, 0)) = \mu(q_0) = q'_0$, the automaton \mathcal{A}' satisfies Property (3).

Properties 4 and 5 are now considered. By Lemma 2.4, since \mathcal{A}' is minimal, there is at most one state in $Q_{\infty, \mathcal{A}'}$ and in $Q_{\emptyset, \mathcal{A}'}$. Furthermore $\delta(q_0, \star\star)$ belongs to $Q_{\emptyset, \mathcal{A}'}$, hence $Q_{\emptyset, \mathcal{A}'}$ is not empty. Hence \mathcal{A}' satisfies Properties 4 and 5.

Property (6) is now considered. Let $0 < a < b$, it is now proven that $\delta'(q'_0, a) \neq q'_0$. By Property (6), $\delta(q_0, a) \neq q_0$. By Lemma 1.8, there exists $i \in \mathbb{N}^{>0}$ such that $\delta(q_0, a^i)$ is recurrent. Let $q = \delta(q_0, a^i)$. By Property (8), q is either q_0 , $q_{\emptyset, \mathcal{A}}$ or $q_{\infty, \mathcal{A}}$. The three cases are considered separately.

- It is first assumed that $q = q_0$, in this case, all of the states $\delta(q_0, a^i)$ are in the same strongly connected component. And since $\delta(q_0, a) \neq q_0$, Property (8) implies that $\delta(q_0, a)$ is either $q_{\emptyset, \mathcal{A}}$ or $q_{\infty, \mathcal{A}}$. The case where $\delta(q_0, a)$ is $q_{\emptyset, \mathcal{A}}$ is now considered. The other case is similar. Since $q_{\emptyset, \mathcal{A}}$ and q_0 are in the same strongly connected component, q_0 is accessible from $q_{\emptyset, \mathcal{A}}$. Therefore \mathcal{A} recognizes the empty language. Having \mathcal{A} recognizing the empty language contradicts Property (2).
- It is now assumed that $q = q_{\emptyset, \mathcal{A}}$, the case $q = q_{\infty, \mathcal{A}}$ is similar. By Property (4), since $q = q_{\emptyset, \mathcal{A}}$, $\mu(q) \in Q_{\emptyset, \mathcal{A}'}$. Furthermore, $\mu(q) = \mu(\delta(q_0, a^i)) = \delta'(\mu(q_0), a^i) = \delta'(q'_0, a^i)$. Note that $\delta'(q'_0, a^i) = q_{\infty, \mathcal{A}'} \neq q'_0$. Since $\delta'(q'_0, a^i) \neq q'_0$ it follows that $\delta'(q'_0, a) \neq q'_0$. Therefore, \mathcal{A}' satisfies Property (6).

Property (7) is now considered. Let us assume that $q_{\infty, \mathcal{A}'}$ exists. Let $q' \in Q'$ and $a \in \Sigma_b$ such that $\delta'(q', a) = q_{\emptyset, \mathcal{A}'}$. It must be proven that $q' \notin Q_{I, \mathcal{A}'} \setminus \{q_{\emptyset, \mathcal{A}'}\}$. Since the automaton is minimal, the strongly connected component of $q_{\infty, \mathcal{A}'}$ is $\{q_{\infty, \mathcal{A}'}\}$. By Lemma 1.14, there exists a strongly connected component C in \mathcal{A} such that $\mu(C) = \{q_{\infty, \mathcal{A}'}\}$. Let $q_{\infty, \mathcal{A}}$ be a state of C , since $\mathcal{A}_{q_{\infty, \mathcal{A}}}$ and $\mathcal{A}'_{\mu(q_{\infty, \mathcal{A}})} = \mathcal{A}'_{q_{\infty, \mathcal{A}'}}$ recognizes the same language, and since $q_{\infty, \mathcal{A}'} \in Q_{\infty, \mathcal{A}'}$ then $q_{\infty, \mathcal{A}} \in Q_{\infty, \mathcal{A}}$. Since $q_{\infty, \mathcal{A}}$ belongs to a strongly connected component, it is recurrent. Since $q_{\infty, \mathcal{A}}$ is a recurrent state belonging to $q_{\infty, \mathcal{A}}$, by Property (7), $\delta(q, a) \neq q_{\emptyset, \mathcal{A}}$ for all $q \in Q_{I, \mathcal{A}} \setminus \{q_{\emptyset, \mathcal{A}}\}$ and $a \in \Sigma_b$. By definition of morphism, there exists a state $q \in Q$ such that $\mu(q) = q'$. By Lemma 2.7, since $q' \in Q_{I, \mathcal{A}'}$, $q \in Q_{I, \mathcal{A}}$. Note that $\mu(\delta(q, a)) = \delta'(\mu(q), a) = \delta'(q', a) = q_{\emptyset, \mathcal{A}'} \in Q_{\emptyset, \mathcal{A}'}$. By Lemma 2.7, since $\mu(\delta(q, a)) \in Q_{\emptyset, \mathcal{A}'}$, $\delta(q, a) \in Q_{\emptyset, \mathcal{A}}$. Since $\delta(q, a) \in Q_{\emptyset, \mathcal{A}}$, by Property (4), $\delta(q, a) = q_{\emptyset, \mathcal{A}}$. It implies that $q \notin Q_{I, \mathcal{A}} \setminus \{q_{\emptyset, \mathcal{A}}\}$. By Lemma 2.7, it implies that $q' \notin Q_{I, \mathcal{A}'} \setminus \{q_{\emptyset, \mathcal{A}'}\}$. Therefore, \mathcal{A}' satisfies Property (7).

Finally, Property (8) is now considered. Let q' be a natural recurrent state of \mathcal{A}' . It must be proven that q' is either q_0 , $q_{\emptyset, \mathcal{A}'}$ or $q_{\infty, \mathcal{A}'}$. The state q' is natural, by Lemma 2.7, its antecedents by μ are natural. The state q' is recurrent hence,

by Lemma 1.14, it admits a recurrent antecedent q . By Property (8) applied to \mathcal{A} , either $q = q_0$, $q = q_{\emptyset, \mathcal{A}}$ or $q = q_{\infty, \mathcal{A}}$. The three cases are considered separately.

- The case $q = q_0$ is first considered, in this case, clearly, $q' = q'_0$.
- The case $q = q_{\emptyset, \mathcal{A}}$ is now considered. As proven above, it implies that $\mu(q) = q_{\emptyset, \mathcal{A}'}$.
- The case where $q = q_{\infty, \mathcal{A}}$ is similar to the preceding case.

Therefore, \mathcal{A}' satisfies Property (8). \square

It is now proven that the set $\mathbb{A}_{\mathbb{R} \geq 0, s}$ satisfies Property (4) of Proposition 3.3.

Lemma 5.11. *The automata of $\mathbb{A}_{\mathbb{R} \geq 0, s}$ recognize simple sets.*

In order to prove this lemma, another lemma is first introduced. It implies that the set R recognized by an automaton $\mathcal{A} \in \mathbb{A}_{\mathbb{R} \geq 0, s}$ with n states is such that $[0, b^n)$ is either a subset of R or is disjoint from R .

Lemma 5.12. *Let $\mathcal{A} \in \mathbb{A}_{\mathbb{R} \geq 0, s}$ be an automaton with n states recognizing a set R . If \mathcal{A} contains a state $q_{\infty, \mathcal{A}}$, as in Definition 5.7, then $(b^{n-1}, \infty) \subseteq R$, otherwise $(b^{n-1}, \infty) \cap R = \emptyset$.*

Proof. Two cases must be considered, depending on whether the state $q_{\infty, \mathcal{A}}$ exists. Let us assume that the state $q_{\infty, \mathcal{A}}$ exists, the other case is similar. Let $x > b^{n-1}$ and let $0^c w_I \star w_F$, be one of its encoding in base b , with $c \in \mathbb{N}$ and $w_I[0] \neq 0$. Let us prove that \mathcal{A} accepts $0^c w_I \star w_F$.

Note that since $x > b^{n-1}$, it implies that the length of w_I is at least n . For $i \leq |w_I|$, let $q_i^I = \delta(q_0, 0^c w_I [< i])$. By Lemma 1.8, there exists $0 \leq i' < i \leq n$ such that $q_{i'}^I$ is recurrent. By Property (8) of Definition 5.7, the only natural recurrent states of \mathcal{A} are $q_{\infty, \mathcal{A}}$, $q_{\emptyset, \mathcal{A}}$ and q_0 . By Properties 6 and 7, $q_i^I \neq q_0$. Since $q_{\infty, \mathcal{A}}$ exists, by Property (7) then $q_i^I \neq q_{\emptyset, \mathcal{A}}$. Since q_i^I is either $q_{\infty, \mathcal{A}}$, $q_{\emptyset, \mathcal{A}}$, or q_0 , since $q_i^I \neq q_{\emptyset, \mathcal{A}}$ and since $q_i^I \neq q_0$ $q_i^I = q_{\infty, \mathcal{A}}$. It follows that $\delta(q_i^I, w_I) = q_{\infty, \mathcal{A}}$, and then, for all $j \in \mathbb{N}$, $\delta(q_i^I, 0^c w_I \star w_F [< j]) = q_{[0, 1], \mathcal{A}}$. Therefore \mathcal{A} accepts $0^c w_I \star w_F$. \square

Example 5.11 is now proven

Proof of Example 5.11. Let $\mathcal{A} \in \mathbb{A}_{\mathbb{R} \geq 0, s}$ with n states and let $R = [\mathcal{A}]_b^{\mathbb{R}}$. Let us prove that R is simple. By Lemma 5.12, it suffices to prove that $R \cap [0, b^{n-1})$ is simple. In order to do this, it suffices to prove that R_i is simple for all $i \in [b^{n-1} - 1]$.

By Property (1) of Definition 5.7, $\mathcal{A}_{\delta(q_0, w \star)} \in \mathbb{A}_{F, S}$ for all $w \in \Sigma_b^*$. Note that the difference between R_i and $R'_i = \bigcup_{j \in \mathbb{N}} [\mathcal{A}_{\delta(q_0, 0^j w \star)}]_b^{\mathbb{R}}$ is a subset of $\{0, 1\}$. Therefore, it suffices to prove that R'_i is simple. Note that, since $\delta(q_0, 0) = q_0$, all $\mathcal{A}_{\delta(q_0, 0^j w \star)}$ are equals. Therefore the infinite union R'_i of simple sets is a simple set. That is, R'_i is a simple set. \square

Proposition 3.3 is now proven. Its proof is similar to the proof of Theorem 4.4.

Proof of Proposition 3.3. It suffices to use Proposition 3.3 with \mathbb{A} being the set of automata $\mathbb{A}_{R \geq 0, S}$, \mathbb{A}' being the set of RNA, and Lemmas 5.8, 5.9, 5.10 and 5.11. \square

Theorem 5.5 is now proven.

Proof of Theorem 5.5. The algorithm to decide whether a minimal weak Büchi RNA \mathcal{A} recognizes a simple set is now given. The algorithm first checks whether Property (2) of Definition 5.7 holds. If it does not hold, then the automaton recognizes the empty language or $\Sigma_b^* \star \Sigma_b^\omega$, in which cases the algorithm accepts. If Property (2) holds, then the automaton recognizes a non empty strict subset of $\Sigma_b^* \star \Sigma_b^\omega$ and therefore the problem is reduced to the problem considered in Proposition 6.2. It thus suffices to apply the algorithm of Proposition 6.2 to \mathcal{A} and to return the result of this algorithm. \square

The algorithm of Theorem 5.5 takes as input a RNA and runs in time $O(nb)$. It should be noted that it is not known whether it is decidable in time $O(nb)$ whether an automaton is a RNA. However, as for the algorithm of Theorem 4.4, if the algorithm of Theorem 5.5 is applied to a weak Büchi automaton which is not a real automaton, the algorithm returns no false positive. An example of false negative is now given. Let L be the language described by the regular expression:

$$(00)^* (01 + 2) \Sigma_3^* \star \Sigma_3^\omega.$$

This language is recognized by the automaton of Figure 6. Note that $[L]_b^R = [1, \infty)$, which is a simple set. However, since Property (3) is not satisfied by the automaton of Figure 6, the algorithm of Theorem 5.5 does not accept this automaton.

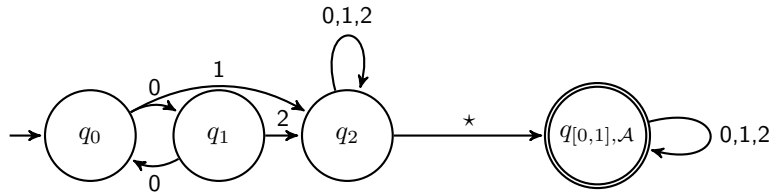


Figure 6: An automaton which recognizes $[1, \infty)$ and is refused by the algorithm of Theorem 5.5.

6 From automata to simple set

In this section, it is explained, given a weak Büchi automaton recognizing a simple set R , how to compute a first-order formula which defines R . The exact theorem is now stated.

Theorem 6.1. *Let \mathcal{A} be a minimal RNA with n states, over the alphabet Σ_b , which recognizes a simple set. There exists two formulas which define $[\mathcal{A}]_b^R$:*

- a $\Sigma_1[\mathbb{R}; +, <, 1]$ -formula computable in time $O(n^2 b \log(nb))$ and
- a $\Sigma_2[\mathbb{R}; +, <, 1]$ -formula computable in time $O(nb \log(nb))$.

In order to prove this theorem, a more general proposition is introduced. This proposition shows that the condition that \mathcal{A} is a RNA can be replaced by the condition $\mathcal{A} \in \mathbb{A}_{R \geq 0, s}$.

Proposition 6.2. *Let $\mathcal{A} \in \mathbb{A}_{F, S}$ be a minimal automaton. There exists two formulas which define $[\mathcal{A}]_b^R$:*

- a $\Sigma_1[\mathbb{R}; +, <, 1]$ -formula computable in time $O(n^2 b \log(nb))$ and
- a $\Sigma_2[\mathbb{R}; +, <, 1]$ -formula computable in time $O(nb \log(nb))$.

In order to prove this proposition, a technical lemma is first introduced.

Lemma 6.3. *Let $A \in \mathbb{A}_{R \geq 0, s}$ be minimal with n states, $w_I \in \Sigma_b^*$, $w_F \in \Sigma_b^\omega$ and $\mathcal{Q} \subseteq Q$ containing exactly one state of each strongly connected component. Then, there exists $s \in [n]$ such that $\delta(q, w_I \star (w_F [< s])) \in \mathcal{Q}$.*

Proof. Let $q_i^F = \delta(q_0, w_I \star (w_F [< i]))$ for any $i \in \mathbb{N}$. By Lemma 1.8, there exists $0 \leq i < i' \leq n$ such that q_i^F is recurrent. Let C be the strongly connected component of q_i^F . By Property (8) of Definition 5.7, there are three kinds of strongly connected components in the fractional part of a minimal automaton of $\mathbb{A}_{R \geq 0, s}$: the singleton $\{q_{\emptyset, \mathcal{A}}\}$, the singleton $\{q_{[0,1], \mathcal{A}}\}$, and the cycles. Three cases must be considered depending on the kind of strongly connected components that is C . In the two first cases, if q_i^F is $q_{\emptyset, \mathcal{A}}$ or $q_{[0,1], \mathcal{A}}$ then $q_i^F \in \mathcal{Q}$ and it suffices to take $s = i$. Otherwise, if C is a cycle, then $\{q_j^F \mid i \leq j < i'\} = C$, therefore there exists an integer $i \leq s < i'$ such that $q_s^F \in C$. \square

Proposition 6.2 can now be proven.

Proof. Let $R = [\mathcal{A}]_b^R$. Since the automaton \mathcal{A} belongs to $\mathbb{A}_{R \geq 0, s}$, the notations of Definition 5.7 can now be used, and therefore the notations of Definition 4.5 can also be used. Recall that all strongly connected components are cycles, apart from $\{q_0\}$, $\{q_{\infty, \mathcal{A}}\}$, $\{q_{[0,1], \mathcal{A}}\}$ and $\{q_{\emptyset, \mathcal{A}}\}$. Furthermore, for each $q \in Q_{F, \mathcal{A}}$ in a cycle C , the digit s_q is the only one such that $\delta(q, s_q) \in C$. In this proof, it is assumed that each state has an integer index between 0 and $n - 1$. For q a state with index i and p a variable, the atomic first-order formula $p \doteq q$ is an

abbreviation for $p \doteq i$. Let $\mathcal{Q} \subseteq Q$ be a set as in Lemma 6.3. Note that the states $q_{\emptyset, \mathcal{A}}$, $q_{[0,1], \mathcal{A}}$ and $q_{\infty, \mathcal{A}}$ all belong to \mathcal{Q} if they exist.

It is first explained how to compute an existential formula $\phi(x)$ which defines R in time $O(n^2 \log(n) b \log(b))$. As seen in Lemma 5.12, the formula which defines R is either $\phi(x) \vee x \geq b^{n-1}$ if $q_{\infty, \mathcal{A}}$ exists, or $\phi(x)$ otherwise. At the end of the proof, it is explained how to decrease the time by adding universal quantifiers.

For any real $0 \leq x < b^{n-1}$, let $x_I \in \mathbb{N}$ and $x^F \in [0, 1]$ be numbers such that $x = x_I + x_F$. If $x \in \mathbb{N}^{>0}$, then the pair (x_I, x_F) is either $(\lfloor x_I \rfloor, x - \lfloor x_I \rfloor)$ or $(x - 1, 1)$. Otherwise, the pair (x_I, x_F) is $(\lfloor x_I \rfloor, x - \lfloor x_I \rfloor)$. Let $w_{x_I} \in \Sigma_b^n$ be an encoding in base b of x_I of length n . Since $x_I < b^{n-1}$, such an encoding exists. Let $w_{x_F} \in \Sigma_b^\omega$ be an encoding in base b of x_F . Let $q_{x,i}^I = \delta(q_0, w_{x_I} \prec i)$ for all $i \in [n]$ and $q_{x,i}^F = \delta(q_0, w_{x_I} \star (w_{x_F} \prec i))$ for all $i \in \mathbb{N}$.

The formula $\phi(x)$ which defines $[\mathcal{A}_R]_b^R$ is defined as the conjunction of two subformulas. Intuitively, the first formula, $\phi_I(p^I, x_I)$ considers the run on w_I and the second formula, $\phi_F(p^F, x_F)$, considers the run on w_F .

Let us assume that there exists a formula $\phi_I(p^I, x_I)$, of size $O(n^2 b \log(nb))$, such that, if $x_{nI} < b^{n-1}$, the formula holds if and only if p^I is the index of w_{x_F} and if $x_F \in \mathbb{N}$. Let us assume that there exists a formula $\phi_F(p^F, x_F)$, of size $O(n^2 b \log(nb))$, which accepts $x_F \in [0, 1]$ if and only if \mathcal{A}_{p^I} accepts an encoding of x_F . Then it suffices to take $\phi(x)$ to be the formula:

$$\begin{aligned} \phi(x) = \exists x_I, x_F, p^I, p^F. & x \doteq x_I + x_F \wedge x_I < b^{n-1} \wedge x_F \in [0, 1] \\ & \wedge \bigvee_{q \in Q_{I, \mathcal{A}}} (p^I \doteq q \wedge p^F \doteq \delta(q, \star)) \\ & \wedge \phi_I(p^I, x_I) \wedge \phi_F(p^F, x_F). \end{aligned}$$

The formula $\phi_I(p^I, x_I)$ is now defined. A sequence $(x_{iI})_{i \in [n]}$ of variables is existentially quantified in this formula. The variable x_{iI} is intended to be interpreted by the value $\lfloor \frac{x_I}{b^{n-i}} \rfloor$, its encoding in base b is $w_{x_I} \prec i$. A sequence $(p_i^I)_{i \in [n]}$ of variables is existentially quantified. They are used to encode the n steps of the run of \mathcal{A} on w_I . More precisely, the variable p_i^I is meant to be interpreted by the index of q_i^I . Note that if $x_{0I} = 0$, since $x_{i+1I} = bx_{iI} + w_x[i]$ for $i \in [n-1]$, an easy induction shows that $x_{iI} \in \mathbb{N}$ for all $0 \leq i \leq n$.

Let us assume that there exists a $\Sigma_0[\mathbb{R}; +, <, 1]$ -formula $\psi^I(p_i^I, x_{iI}, p_{i+1}^I, x_{i+1I})$, of size $O(nb \log(nb))$, which asserts that if p_i is the index of a state q , and if $x_{i+1I} = bx_{iI} + a$ for some $a \in \Sigma_b$, then p_{i+1} is the index of $\delta(q, a)$. The formula $\phi_I(p^I, x_I)$ can then be taken to be the $\Sigma_1[\mathbb{R}; +, <, 1]$ -formula of length $O(n^2 b \log(nb))$:

$$\begin{aligned} \phi_I(p^I, x_I) = \exists (p_i^I, x_{iI})_{0 \leq i \leq n}. & x_{0I} \doteq 0 \wedge p_0^I \doteq q_0 \\ & \wedge \bigwedge_{i=0}^{n-1} \psi^I(p_i^I, x_{iI}, p_{i+1}^I, x_{i+1I}) \\ & \wedge x_{nI} \doteq x_I \wedge p_n^I \doteq p^I. \end{aligned}$$

Formally, the formula $\psi_I(p_i^I, x_{iI}, p_{i+1}^I, x_{i+1I})$ can be taken to be:

$$\psi_I \left(\begin{array}{c} p_i^I, x_{iI}, \\ p_{i+1}^I, x_{i+1I} \end{array} \right) = \bigwedge_{q \in Q_{I,A}} \bigwedge_{a \in \Sigma_b} \left[\begin{array}{c} (x_{i+1I} \doteq bx_{iI} + a \wedge p_i^I \doteq q) \implies \\ p_{i+1}^I \doteq \delta(q, a) \end{array} \right].$$

No notations introduced during the construction of the formula $\phi_I(p^I, x_I)$, are used in the remaining of the proof.

The formula $\phi_F(p^F, x_F)$ is now defined. Let s be the smallest integer such that $q_{x,s}^F \in \mathcal{Q}$, by Lemma 6.3, such an integer exists. The formula $\phi_F(p^F, x_F)$ is the conjunction of two subformulas. The first formula, $\phi_{1F}(p, y, p^F, x_F)$, considers the run of \mathcal{A} on the first s letters of w_{x_F} . The second formula $\phi_{2F}(p, y)$ considers the end of the run, on $w_x[\geq s]$, beginning at the state $q_{x,s} \in \mathcal{Q}$.

Two variables p and y are existentially quantified. They are meant to be interpreted by p_s^F and $w_{x_F}[\geq s]$ respectively. Assume that there is a $\Sigma_1[\mathbb{R}; +, <, 1]$ -formula, $\phi_{1F}(p, y, p^F, x_F)$, of length $O(n^2 b \log(b))$, which states that, given p^F , x_F 's, the variables p, y are interpreted as stated above. Let us also assume that there exists a $\Sigma_0[\mathbb{R}; +, <, 1]$ -formula $\phi_{2F}(p, y)$, of length $O(n^2 \log(b))$, which states that \mathcal{A}_q accepts an encoding of y , where p is the index of the state q . Then the formula ϕ_F can be taken to be the $\Sigma_1[\mathbb{R}; +, <, 1]$ -formula of length $O(n^2 b \log(nb))$:

$$\phi_F(p^F, x_F) = \exists p, y. \phi_{1F}(p, y, p^F, x_F) \wedge \phi_{2F}(p, y).$$

The formula, $\phi_{1F}(p, y, p^F, x_F)$ is now constructed. As in the construction of $\phi_F(p^F, x_F)$, two sequences of variables are existentially quantified to encode some suffix of w_F and to encode a part of the run of the automaton on w . The sequence $(x_{iF})_{i \in [n]}$ is existentially quantified. The variable x_{iF} is meant to be interpreted by $[w_{x_F}[\geq i]]_b^{\mathbb{R}}$. It represents the real that must be read at the i -th step of the run after the \star . It follows that $x_{0F} = x_F$, $x_{sF} = y$ and that x_{iF} is equal to $bx_{i-1} - w_{x_F}[i]$. The sequence $(p_i^F)_{i \in [n]}$ of variables is existentially quantified. It is used to encode the first $(n+1)$ steps of the run after the \star part of the run. More precisely, the variable p_i^F is meant to be interpreted by the indexes of q_i^F . A third sequence of variables, $(s_i)_{i \in [n]}$, is existentially quantified. The variable s_i is meant to be interpreted by 0 if $i < s$ and 1 otherwise. Note that $s_0 = 0$ as $s \geq 0$. Those variables allows to know the value of s .

Let $\psi^F(p, y, p_i^F, x_{iF}, s_i, p_{i+1}^F, x_{i+1F}, s_{i+1})$ be $\Sigma_0[\mathbb{R}; +, <, 1]$ -formula of length $O(nb \log(nb))$ which states that, given p_i^F, x_{iF}, s_i , the variables p_{i+1}^F, x_{i+1F} and s_{i+1} are correctly interpreted, and furthermore if $i = s$ – that is if $p_i^F \in \mathcal{Q}$ and for all $j < i$, $p_j^F \notin \mathcal{Q}$ – then the variables p and y are correctly interpreted. The formula $\phi_{1F}(p, y, p^F, x_F)$ can then be expressed as the $\Sigma_1[\mathbb{R}; +, <, 1]$ -formula of length $O(n^2 b \log(nb))$:

$$\phi_{1F} \left(\begin{array}{c} p, y, \\ p^F, x_F \end{array} \right) = \begin{array}{c} \exists (p_i^F, x_{iF}, s_i)_{i \in [n]}. p^F \doteq p_0^F \wedge x_F \doteq x_{0F} \wedge \\ \bigwedge_{i=0}^{n-1} \psi^F \left(\begin{array}{c} p, y, p_i^F, x_{iF}, \\ s_i, p_{i+1}^F, x_{i+1F}, s_{i+1} \end{array} \right). \end{array}$$

The formula $\psi_F(p, y, p_i^F, x_{iF}, s_i, p_{i+1}^F, x_{i+1F}, s_{i+1})$ is now given. It is the $\Sigma_0[\mathbb{R}; +, <, 1]$ -formula of length $O(nb \log(nb))$:

$$\psi_F \left(\begin{matrix} p_i^F, & x_{iF}, & s_i, & p \\ p_{i+1}^F, & x_{i+1F}, & s_{i+1} & y \end{matrix} \right) = \left\{ (s_{i+1} \doteq 1) \iff \left(s_i \doteq 1 \vee \bigvee_{q \in \mathcal{Q}} p_i^F \doteq q \right) \right\} \wedge \left\{ \bigvee_{q \in \mathcal{Q}} \bigvee_{a=0}^{b-1} \left[\begin{matrix} p_i^F \doteq q & \wedge & x_{iF} \in [\frac{a}{b}, \frac{a+1}{b}] \\ p_{i+1}^F \doteq \delta(q, a) \wedge x_{i+1F} \doteq b(x_{iF} - \frac{a}{b}) \end{matrix} \right] \wedge \right\} \wedge \left\{ [s_{i+1} \doteq 1 \wedge s_i \doteq 0] \implies [p \doteq p_i^F \wedge y \doteq x_{iF}] \right\}.$$

The formula $\phi_{2F}(t, y)$ is now constructed. It is a disjunction, which states that there exists $q \in C$, such that $q_{x,s} = q$, and such that $\mathcal{A}_{q_{x,s}}$ accepts an encoding of $y = x_{sF}$. By definition of s , $q_{x,s} \in \mathcal{Q}$. Let us assume that, for each $q \in \mathcal{Q}$ in a strongly connected component C , there exists a $\Sigma_0[\mathbb{R}; +, <, 1]$ -formula $\xi_q(y)$ of length $O(\log(b)|C|)$, where $|C|$ is the cardinal of C , which states that \mathcal{A}_q accepts an encoding of y . Then, $\phi_{2F}(p, y)$ can be taken as the $\Sigma_0[\mathbb{R}; +, <, 1]$ -formula of length $O(n \log(nb))$:

$$\phi_{2F}(p, y) = \bigvee_{q \in \mathcal{Q}} p = q \wedge \xi_q(y).$$

Let us now construct the formula $\xi_q(y)$. Trivially, $\xi_{q_{[0,1],\mathcal{A}}}(y)$ can be taken to be True and $\xi_{q_{\emptyset,\mathcal{A}}}(y)$ can be taken to be False. They are constant size formula. It remains to construct the formulas $\xi_q(y)$ for q in a cycle C . Let s_q be the value s_q defined as in the proof of Lemma 4.10, for the automaton \mathcal{A}_q . As shown in the proof of Lemma 4.10, the length of s_q is $O(\log(b)|C|)$ and the automaton \mathcal{A}_q recognizes a set, which is a union of $[0, s_q]$, $\{s_q\}$, and $(s_q, 1]$, and which is defined by a formula ξ_q of length $O(\log(b)|C|)$.

It is now explained how to transform the $\Sigma_1[\mathbb{R}; +, <, 1]$ -formula $\phi(x)$ of length $O(n^2 b \log(nb))$ into an equivalent $\Sigma_2[\mathbb{R}; +, <, 1]$ -formula of length $O(nb \log(nb))$. Let us assume that there exists $\phi_1'^{[0,1]}(p, y, p^F, x_F)$ and $\phi'^{\mathbb{N}}(p^I, x_I)$, two $\Sigma_2[\mathbb{R}; +, <, 1]$ -formulas of length $O(nb \log(nb))$, equivalent to $\phi_{1F}(p, y, p^F, x_F)$ and to $\phi_I(p^I, x_I)$ respectively. It thus suffices to replace the two formulas $\phi_{1F}(p, y, p^F, x_F)$ and $\phi_I(p^I, x_I)$ in $\phi(x)$ by their equivalent smaller formulas.

In order to construct a $\Sigma_2[\mathbb{R}; +, <, 1]$ -formula of length $O(nb \log(nb))$ equivalent to $\phi_{1F}(p, y, p^F, x_F)$ or to $\phi_I(p^I, x_I)$, it suffices to replace their last conjunctions by universal quantifications. The formula $\phi_{1F}(p, y, p^F, x_F)$ is equivalent to the following $\Sigma_2[\mathbb{R}; +, <, 1]$ -formula of length $O(nb \log(nb))$:

$$\phi_{1F} \left(\begin{matrix} p, y, \\ p^F, x_F \end{matrix} \right) = \bigvee_{\substack{\rho, \xi, \gamma, \\ \rho', \xi', \gamma'}} \left\{ \begin{matrix} \exists (p_i^F, x_{iF}, s_i)_{i \in [n]}. p^F \doteq p_0^F \wedge x_F \doteq x_{0F} \wedge \\ \left[\bigvee_{i=0}^{n-1} \left(\begin{matrix} \rho = p_i^F & \wedge & \xi = x_{iF} & \wedge & \gamma = s_i \\ \rho' = p_{i+1}^F & \wedge & \xi' = x_{i+1F} & \wedge & \gamma' = s_{i+1} \end{matrix} \right) \right] \wedge \\ \implies \psi_F \left(\begin{matrix} \rho, \xi, \gamma, t \\ \rho', \xi', \gamma', y \end{matrix} \right) \end{matrix} \right\}.$$

Similarly, the formula $\phi_I(p_0^I, x_{0I})$ is equivalent to the $\Sigma_2[\mathbb{R}; +, <, 0]$ -formula of

length $O(nb \log(nb))$:

$$\phi_I(p^I, x_I) = \exists (p^I, x_{iI})_{0 \leq i \leq n} \cdot \forall \rho, \xi, \rho', \xi' \cdot \left\{ \begin{array}{l} x_{0I} \doteq 0 \wedge p_0^I \doteq q_0 \\ \wedge x_{nI} \doteq x_I \wedge p_n^I \doteq p^I \wedge \\ \left[\bigvee_{i=0}^{n-1} \left(\rho \doteq p_i^I \wedge \xi \doteq x_{iI} \wedge \right. \right. \\ \left. \left. \rho' \doteq p_{i+1}^I \wedge \xi' \doteq x_{i+1I} \right) \right] \\ \implies \psi_F(\rho, \xi, \rho', \xi') \end{array} \right\}.$$

□

Theorem 6.1 can now be proven.

Proof of Theorem 6.1. The algorithm is exactly the same than the algorithm of Proposition 6.2. It suffices to prove that the algorithm of Proposition 6.2 can be applied to \mathcal{A}' , that is, that $\mathcal{A}' \in \mathbb{A}_{F,S}$.

Let $L = L_\omega(\mathcal{A})$ and $R = [L]_b^R$. Since \mathcal{A} is fractional, then L is also fractional, hence $L = L_\omega(R)$. By Lemma 4.3, L is also recognized by \mathcal{A}_R as in Definition 4.1. By Lemma 4.8, $\mathcal{A}_R \in \mathbb{A}_{F,S}$ and by Lemma 4.9, its minimal quotient \mathcal{A}'' belongs to $\mathbb{A}_{F,S}$. Since \mathcal{A} and \mathcal{A}_R recognizes the same language, \mathcal{A}'' is also the minimal quotient of \mathcal{A} , therefore $\mathcal{A}'' = \mathcal{A}'$ and $\mathcal{A}' \in \mathbb{A}_{F,S}$. □

7 Conclusion

In this paper, we proved that it is decidable in linear time whether a weak Büchi Real Number Automaton \mathcal{A} reading a set of real number R recognizes a finite union of intervals. It is proved that a quasi-linear sized existential-universal formula defining R exists. And that a quasi-quadratic existential formula defining R also exists.

The theorems of this paper lead us to consider two natural generalization. We intend to adapt the algorithm of this paper to similar problems for automata reading vectors of reals instead of automata reading reals. We also would like to solve a similar problem, deciding whether an RNA accepts a $\text{FO}[\mathbb{R}, \mathbb{Z}; +, <]$ -definable set of reals. Solving this problem would also solve the problem of deciding whether an automaton reading natural number, beginning by the most-significant digit, recognizes an ultimately-periodic set. Similar problems has already been studied, see e.g [1, 10] and seems to be difficult.

The author thanks Bernard Boigelot, for a discussion about the algorithm of Theorem 6.1, which led to a decrease of the computation time.

References

- [1] Boris Alexeev. Minimal dfas for testing divisibility. *CoRR*, cs.CC/0309052, 2003.
- [2] Bernard Boigelot, Julien Brusten, and Véronique Bruyère. On the sets of real numbers recognized by finite automata in multiple bases. *Logical Methods in Computer Science*, 6(1), 2010.

- [3] Bernard Boigelot, Julien Brusten, and Jérôme Leroux. A Generalization of Semenov's Theorem to Automata over Real Numbers. In Renate A. Schmidt, editor, *Automated Deduction, 22nd International Conference, CADE 2009*, volume 5663 of *Lecture Notes in Computer Science*, pages 469–484, Montréal, Canada, August 2009. Springer Berlin.
- [4] Bernard Boigelot, Sébastien Jodogne, and Pierre Wolper. An effective decision procedure for linear arithmetic over the integers and reals. *ACM Trans. Comput. Logic*, 6(3):614–633, July 2005.
- [5] Véronique Bruyère, Georges Hansel, Christian Michaux, and Roger Villemaire. Logic and p-recognizable sets of integers. *Bull. Belg. Math. Soc.*, 1:191–238, 1994.
- [6] Alan Cobham. On the base-dependence of sets of numbers recognizable by finite automata. *Mathematical systems theory*, 3(2):186–192, 1969.
- [7] Jeanne Ferrante and Charles Rackoff. A decision procedure for the first order theory of real addition with order. *SIAM J. Comput.*, 4(1):69–76, 1975.
- [8] Godfrey Harold Hardy and Edward Maitland Wright. An introduction to the theory of numbers, 1960. Autres tirages avec corrections : 1962, 1965, 1968, 1971, 1975.
- [9] Juha Honkala. A decision method for the recognizability of sets defined by number systems. *ITA*, 20(4):395–403, 1986.
- [10] Anne Lacroix, Narad Rampersad, Michel Rigo, and Élise Vandomme. Syntactic complexity of ultimately periodic sets of integers and application to a decision procedure. *Fundam. Inform.*, 116(1-4):175–187, 2012.
- [11] Jérôme Leroux. Least significant digit first Presburger automata. *CoRR*, abs/cs/0612037, 2006.
- [12] Christof Löding. Efficient minimization of deterministic weak omega automata, 2001.
- [13] Victor Marsault and Jacques Sakarovitch. Ultimate periodicity of b-recognisable sets: A quasilinear procedure. In Marie-Pierre Béal and Olivier Carton, editors, *Developments in Language Theory*, volume 7907 of *Lecture Notes in Computer Science*, pages 362–373. Springer, 2013.
- [14] Andrei A. Muchnik. The definable criterion for definability in Presburger arithmetic and its applications. *Theor. Comput. Sci.*, 290(3):1433–1444, 2003.
- [15] A. L. Semenov. The presburger nature of predicates that are regular in two number systems. *Siberian Math. J.*, 18:289,299, 1977.

- [16] Robert Tarjan. Depth-first search and linear graph algorithms. *SIAM Journal on Computing*, 1(2):146–160, 1972.
- [17] Volker Weispfenning. Mixed real-integer linear quantifier elimination, 1999.
- [18] Pierre Wolper and Bernard Boigelot. On the construction of automata from linear arithmetic constraints. In Susanne Graf and Michael Schwartzbach, editors, *Tools and Algorithms for the Construction and Analysis of Systems*, volume 1785 of *Lecture Notes in Computer Science*, pages 1–19. Springer Berlin Heidelberg, 2000.

Index

- $q_{\emptyset, \mathcal{A}}$, 13
- $q_{\infty, \mathcal{A}}$, 13
- $[L]_b^{\mathbb{R}}$, 5
- $[\mathcal{A}]_b^{\mathbb{R}}$, 12
- $q_{[0,1], \mathcal{A}}$, 13
- $[\cdot]_b$, 4
- Deterministic Büchi automaton, 6
- $\mathbb{A}_{F, \mathcal{S}}$, 23
- $\mathbb{A}_{\mathbb{R}^{\geq 0}, \mathcal{S}}$, 28
- \mathcal{A}_q for \mathcal{A} an automaton and q a state., 7
- $L_{\omega}(\mathcal{A})$, 7
- Accessible, 7
- \mathcal{A}_R - for $R \subseteq \mathbb{R}^{\geq 0}$, 26
- \mathcal{A}_R and \mathcal{A}'_R - for a simple set $R \subseteq [0, 1]$, 20
- FNA - Fractional Number Automaton, 12
- Fractional Number Automaton, 12
- Fractional part of a word of $\Sigma_b^* \star \Sigma_b^{\omega}$, 4
- Leaf, 7
- Length of formulas, 11
- Morphism of automata, 8
- Morphism of automata, 8
- \mathbb{N} , 4
- $[n]$ for $n \in \mathbb{N}$., 4
- Natural part of a word of $\Sigma_b^* \star \Sigma_b^{\omega}$, 4
- Q_q for Q a set of states and q a state., 7
- \mathbb{Q} , 4
- $Q_{[0,1], \mathcal{A}}$, 13
- $Q_{F, \mathcal{A}}$, 13
- $Q_{\infty, \mathcal{A}}$, 13
- $Q_{I, \mathcal{A}}$, 13
- $Q_{\emptyset, \mathcal{A}}$, 13
- Quotient of automata, 8
- \mathbb{R} , 4
- $\lfloor r \rfloor$, 4
- Real Number Automaton, 12
- Recognizable language, 7
- Recurrent state, 7
- Rejecting state, 6
- RNA- Real Number Automaton, 12
- Run, 7
- Run of an automaton, 6
- Saturated, 5
- Simple sets, 11
- Strongly connected components, 7
- Threshold of a simple set., 11
- Transient state, 7
- $u_{i,j,k}$, 19
- $v_{i,j,k}$, 19
- Weakly recognizable language, 8
- \mathbb{Z} , 4